BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Claims 21-30 and 32-42 are pending in the application. Claims 32-34 and 38-42 are withdrawn as being drawn to non-elected inventions. Claims 21-30 and 35-37 are under consideration. Claims 21 and 30 have been amended. Support for these amendments can be found in the specification, for example, at page 21, lines 9-14 and page 22, line 23 through page 23, line 4. These amendments further clarify the intended subject matter of the claimed invention. Entry of these amendments is respectfully requested. Applicants reserve the right to prosecute non-elected subject matter in subsequent divisional applications.

Rejoinder

Applicants reiterate their request that claims 32-34 and 41-42, drawn to methods of using the polynucleotides, and claims 38-40, drawn to methods of using the polypeptides, be rejoined per the Commissioner's Notice in the Official Gazette of March 26, 1996, entitled "Guidance on Treatment of Product and Process Claims in light of *In re Ochiai, In re Brouwer* and 35 U.S.C. § 103(b)" which sets forth the rules, upon allowance of product claims, for rejoinder of process claims covering the same scope of products. Applicants request that claims 32-34 and 41-42 be rejoined and examined upon allowance of any of the claims drawn to the polynucleotides of Group A and that claims 39 and 40 be rejoined and examined upon allowance of any of the claims drawn to the polypeptides of Group A.

The Final Rejection

Claims 21-30 and 35-37 stand rejected under 35 U.S.C. §§ 101 and 112, first paragraph, based on the allegation that the claimed invention lacks patentable utility. The rejection alleges in particular that:

- **the claimed invention is not supported by a specific and substantial utility or a well-established utility.**
- **merely identifying that a protein is homologous to the gp120 receptor does not provide sufficient support for a specific and substantial or well-established utility.**

- **absent a disclosure of altered levels or forms of a gene in diseased tissue as compared with the corresponding healthy tissue, the gene is not a disease marker or an appropriate target for drug discovery or toxicology testing.**
- **utilities that require or constitute carrying out further research to identify or reasonably confirm a "real world" context of use are not substantial utilities.**

Claims 21, 23, 26, 27, 28, 30, 35, and 37 stand rejected under 35 U.S.C. § 112, first paragraph, based on the allegation that the specification contains "subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention." The rejection alleges in particular that:

- **it cannot be established that a representative number of species have been disclosed to support the genus claim.**

Claims 21-30 and 35-37 stand rejected under 35 U.S.C. § 112, second paragraph, based on the allegation that the term "naturally occurring" is indefinite. The rejection alleges in particular that:

- **because all of the sequences existing in nature have not been identified, it is not known which sequences would be "naturally occurring" and which would not.**

<u>Issue 1 – Whether the claims meet the utility requirement of 35 U.S.C. § 101</u>

The rejection of claims 21-30 and 35-37 is improper, as the inventions of those claims have a patentable utility as set forth in the instant specification, and/or a utility well known to one of ordinary skill in the art.

The invention at issue comprises polynucleotides expressed in hematopoietic/immune, gastrointestinal, cardiovascular, and reproductive tissues (Specification, e.g., at page 6 and Table 3). The invention also comprises polypeptides encoded by the claimed polynucleotides. The claimed polypeptides are identified in the patent application as human cell surface receptor

proteins, abbreviated as HCSRP. As such, the claimed invention has numerous practical, beneficial uses in toxicology testing, drug development, and the diagnosis of disease, none of which require knowledge of how the polypeptide actually functions.

The similarity of the claimed polypeptide to another polypeptide of known, undisputed utility by itself demonstrates utility beyond the reasonable probability required by law. HCSRP-12 is, in that regard, homologous to non-CD4 glycoprotein gp120 receptor (GENESEQ AAR32188) (Specification, e.g., at Table 2). In particular, SEQ ID NO:12 shares 84% sequence identity with the gp120 receptor (see CLUSTALW alignment attached at Exhibit A).

This is more than enough homology to demonstrate a reasonable probability that the utility of the gp120 receptor can be imputed to the claimed invention. It is well-known that the probability that two unrelated polypeptides share more than 40% sequence homology over 70 amino acid residues is exceedingly small. Brenner *et al.*, Proc. Natl. Acad. Sci. U.S.A. 95:6073-78 (1998). Given homology in excess of 40% over many more than 70 amino acid residues, the probability that the claimed polypeptide is related to the gp120 receptor is, accordingly, very high.

The fact that the claimed polypeptide is a member of the C-type lectin receptor family alone demonstrates utility. Each of the members of this class, regardless of their particular functions, are useful. There is no evidence that any member of this class of polypeptides, let alone a substantial number of them, would not have some patentable utility. It follows that there is a more than substantial likelihood that the claimed polypeptide also has patentable utility, regardless of its actual function. The law has never required a patentee to prove more.

There is, in addition, direct proof of the utility of the claimed invention. Applicants submitted previously the Declarations of Bedilion and Furness describing some of the practical uses of the claimed invention in gene and protein expression monitoring applications as they would have been understood at the time of the patent application.

The Bedilion Declaration describes, in particular, how the claimed expressed polynucleotide can be used in gene expression monitoring applications that were well-known at the time the patent application was filed, and how those applications are useful in developing drugs and monitoring their activity. Dr. Bedilion states that the claimed invention is a useful tool when employed as a highly specific probe in a cDNA microarray:

> Persons skilled in the art would appreciate that cDNA microarrays that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool than cDNA microarrays that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity.

The Patent Examiner does not dispute that the claimed polynucleotide can be used as a probe in cDNA microarrays and used in gene expression monitoring applications. Instead, the Patent Examiner contends that the claimed polynucleotide cannot be useful without precise knowledge of its biological function. But the law never has required knowledge of biological function to prove utility. It is the claimed invention's uses, not its functions, that are the subject of a proper analysis under the utility requirement.

In any event, as demonstrated by the Bedilion Declaration, the person of ordinary skill in the art can achieve beneficial results from the claimed polynucleotide in the absence of any knowledge as to the precise function of the protein encoded by it. The uses of the claimed polynucleotide in gene expression monitoring applications are in fact independent of its precise function.

The Furness Declaration describes, in particular, how the claimed polypeptide can be used in protein expression analysis techniques such as 2-D PAGE gels and western blots. Using the claimed invention with these techniques, persons of ordinary skill in the art can better assess, for example, the potential toxic affect of a drug candidate. (Furness Declaration at ¶ [11]).

The Patent Examiner does not dispute that the claimed polypeptide can be used in 2-D PAGE gels and western blots to perform drug toxicity testing. Instead, the Patent Examiner contends that the claimed polypeptide cannot be useful without precise knowledge of its function. But the law never has required knowledge of biological function to prove utility. It is the claimed invention's uses, not its functions, that are the subject of a proper analysis under the utility requirement.

In any event, as demonstrated by the Furness Declaration, the person of ordinary skill in the art can achieve beneficial results from the claimed polypeptide in the absence of any knowledge as to the precise function of the protein. The uses of the claimed polypeptide for gene

expression monitoring applications including toxicology testing are in fact independent of its precise function.

I. The Applicable Legal Standard

To meet the utility requirement of sections 101 and 112 of the Patent Act, the patent applicant need only show that the claimed invention is "practically useful," *Anderson v. Natta*, 480 F.2d 1392, 1397, 178 USPQ 458 (CCPA 1973) and confers a "specific benefit" on the public. *Brenner v. Manson*, 383 U.S. 519, 534-35, 148 USPQ 689 (1966). As discussed in a recent Court of Appeals for the Federal Circuit case, this threshold is not high:

> An invention is "useful" under section 101 if it is capable of providing some identifiable benefit. See *Brenner v. Manson*, 383 U.S. 519, 534 [148 USPQ 689] (1966); *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 [24 USPQ2d 1401] (Fed. Cir. 1992) ("to violate Section 101 the claimed device must be totally incapable of achieving a useful result"); *Fuller v. Berger*, 120 F. 274, 275 (7th Cir. 1903) (test for utility is whether invention "is incapable of serving any beneficial end").

Juicy Whip Inc. v. Orange Bang Inc., 51 USPQ2d 1700 (Fed. Cir. 1999).

While an asserted utility must be described with specificity, the patent applicant need not demonstrate utility to a certainty. In *Stiftung v. Renishaw PLC*, 945 F.2d 1173, 1180, 20 USPQ2d 1094 (Fed. Cir. 1991), the United States Court of Appeals for the Federal Circuit explained:

> An invention need not be the best or only way to accomplish a certain result, and it need only be useful to some extent and in certain applications: "[T]he fact that an invention has only limited utility and is only operable in certain applications is not grounds for finding lack of utility." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762, 221 USPQ 473, 480 (Fed. Cir. 1984).

The specificity requirement is not, therefore, an onerous one. If the asserted utility is described so that a person of ordinary skill in the art would understand how to use the claimed invention, it is sufficiently specific. *See Standard Oil Co. v. Montedison, S.p.a.*, 212 U.S.P.Q. 327, 343 (3d Cir. 1981). The specificity requirement is met unless the asserted utility amounts to a "nebulous expression" such as "biological activity" or "biological properties" that does not convey meaningful information about the utility of what is being claimed. *Cross v. Iizuka*, 753 F.2d 1040, 1048 (Fed. Cir. 1985).

In addition to conferring a specific benefit on the public, the benefit must also be "substantial." *Brenner*, 383 U.S. at 534. A "substantial" utility is a practical, "real-world" utility. *Nelson v. Bowler*, 626 F.2d 853, 856, 206 USPQ 881 (CCPA 1980).

If persons of ordinary skill in the art would understand that there is a "well-established" utility for the claimed invention, the threshold is met automatically and the applicant need not make any showing to demonstrate utility. Manual of Patent Examination Procedure at § 706.03(a). Only if there is no "well-established" utility for the claimed invention must the applicant demonstrate the practical benefits of the invention. *Id.*

Once the patent applicant identifies a specific utility, the claimed invention is presumed to possess it. *In re Cortright*, 165 F.3d 1353, 1357, 49 USPQ2d 1464 (Fed. Cir. 1999); *In re Brana*, 51 F.3d 1560, 1566; 34 USPQ2d 1436 (Fed. Cir. 1995). In that case, the Patent Office bears the burden of demonstrating that a person of ordinary skill in the art would reasonably doubt that the asserted utility could be achieved by the claimed invention. *Id.* To do so, the Patent Office must provide evidence or sound scientific reasoning. *See In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). If and only if the Patent Office makes such a showing, the burden shifts to the applicant to provide rebuttal evidence that would convince the person of ordinary skill that there is sufficient proof of utility. *Brana*, 51 F.3d at 1566. The applicant need only prove a "substantial likelihood" of utility; certainty is not required. *Brenner*, 383 U.S. at 532.

II. Uses of the claimed polypeptides and polynucleotides for diagnosis of conditions and disorders characterized by expression of HCSRP, for toxicology testing, and for drug discovery are sufficient utilities under 35 U.S.C. §§ 101 and 112, first paragraph

The claimed invention meets all of the necessary requirements for establishing a credible utility under the Patent Law: There are "well-established" uses for the claimed invention known to persons of ordinary skill in the art, and there are specific practical and beneficial uses for the invention disclosed in the patent application's specification. These uses are explained, in detail, in the Bedilion Declaration and the Furness Declaration accompanying this paper. Objective evidence, not considered by the Patent Office, further corroborates the credibility of the asserted utilities.

A. The use of HCSRP for toxicology testing, drug discovery, and disease diagnosis are practical uses that confer "specific benefits" to the public

The claimed invention has specific, substantial, real-world utility by virtue of its use in toxicology testing, drug development and disease diagnosis through gene expression profiling. These uses are explained in detail in the accompanying Bedilion Declaration and Furness Declaration, the substance of which is not rebutted by the Patent Examiner. There is no dispute that the claimed polynucleotide is in fact a useful tool in cDNA microarrays used to perform gene expression analysis and that the claimed polypeptide is a useful tool in two-dimensional polyacrylamide gel electrophoresis ("2-D PAGE") analysis and western blots used to monitor protein expression and assess drug toxicity. These uses are sufficient to establish utilities for the claimed polynucleotide and polypeptide, respectively.

The instant application is a divisional of, and claims priority to, United States Provisional Patent Application Serial No. 60/123,404 filed on March 8, 1999 (hereinafter "the Tang et al. '404 application").

1. The Bedilion Declaration

In his Declaration, Dr. Bedilion explains the many reasons why a person skilled in the art reading the Tang et al. '404 application on March 8, 1999 would have understood that application to disclose the claimed polynucleotide to be useful for a number of gene expression monitoring applications, *e.g.*, as a highly specific probe for the expression of that specific polynucleotide in connection with the development of drugs and the monitoring of the activity of such drugs. (Bedilion Declaration at, e.g., ¶¶ 10-15). Much, but not all, of Dr. Bedilion's explanation concerns the use of the claimed polynucleotide in cDNA microarrays of the type first developed at Stanford University for evaluating the efficacy and toxicity of drugs, as well as for other applications. (Bedilion Declaration, ¶¶ 12 and 15).[1]

---

[1]Dr. Bedilion also explained, for example, why persons skilled in the art would also appreciate, based on the Tang et al. '404 specification, that the claimed polynucleotide would be useful in connection with developing new drugs using technology, such as Northern analysis, that predated by many years the development of the cDNA technology (Bedilion Declaration, ¶ 16).

In connection with his explanations, Dr. Bedilion states that the "Tang et al. '404 specification would have led a person skilled in the art on March 8, 1999 who was using gene expression monitoring in connection with working on developing new drugs for the treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders [a] to conclude that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a highly useful tool, and [b] to request specifically that any cDNA microarray that was being used for such purposes contain the SEQ ID NO:12-encoding polynucleotides" (Bedilion Declaration, ¶ 15). For example, as explained by Dr. Bedilion, "[p]ersons skilled in the art would [have appreciated on March 8, 1999] that a cDNA microarray that contained the SEQ ID NO:12-encoding polynucleotides would be a more useful tool than a cDNA microarray that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity." *Id.*

In support of those statements, Dr. Bedilion provided detailed explanations of how cDNA technology can be used to conduct gene expression monitoring evaluations, with extensive citations to pre-March 8, 1999 publications showing the state of the art on March 8, 1999. (Bedilion Declaration, ¶ ¶ 10-14). While Dr. Bedilion's explanations in paragraph 15 of his Declaration include more than three pages of text and six subparts (a)-(f), he specifically states that his explanations are not "all-inclusive." *Id.* For example, with respect to toxicity evaluations, Dr. Bedilion had earlier explained how persons skilled in the art who were working on drug development on March 8, 1999 (and for several years prior to March 8, 1999) "without any doubt" appreciated that the toxicity (or lack of toxicity) of any proposed drug was "one of the most important criteria to be evaluated in connection with the development of the drug" and how the teachings of the Tang et al. '404 application clearly include using differential gene expression analyses in toxicity studies (Bedilion Declaration, ¶ 10).

Thus, the Bedilion Declaration establishes that persons skilled in the art reading the Tang et al. '404 application at the time it was filed "would have wanted their cDNA microarray to have a [SEQ ID NO:12-encoding polynucleotide probe] because a microarray that contained such a probe (as compared to one that did not) would provide more useful results in the kind of gene

expression monitoring studies using cDNA microarrays that persons skilled in the art have been doing since well prior to March 8, 1999" (Bedilion Declaration, ¶ 15, item (f)). This, by itself, provides more than sufficient reason to compel the conclusion that the Tang et al. '404 application disclosed to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the claimed polynucleotide.

Nowhere does the Patent Examiner address the fact that, as described on p. 33 of the Tang et al. '404 application, the claimed polynucleotides can be used as highly specific probes in, for example, cDNA microarrays – probes that without question can be used to measure both the existence and amount of complementary RNA sequences known to be the expression products of the claimed polynucleotides. The claimed invention is not, in that regard, some random sequence whose value as a probe is speculative or would require further research to determine.

Given the fact that the claimed polynucleotide is known to be expressed, its utility as a measuring and analyzing instrument for expression levels is as indisputable as a scale's utility for measuring weight. This use as a measuring tool, regardless of how the expression level data ultimately would be used by a person of ordinary skill in the art, by itself demonstrates that the claimed invention provides an identifiable, real-world benefit that meets the utility requirement. *Raytheon* v. *Roper*, 724 F.2d 951, (Fed. Cir. 1983) (claimed invention need only meet one of its stated objectives to be useful); *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999) (how the invention works is irrelevant to utility); MPEP § 2107 ("Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific, and unquestionable utility (e.g., they are useful in analyzing compounds)" (emphasis added)).

The Bedilion Declaration shows that a number of pre-March 8, 1999 publications confirm and further establish the utility of cDNA microarrays in a wide range of drug development gene expression monitoring applications at the time the Tang et al. '404 application was filed (Bedilion Declaration ¶¶ 10-14; Bedilion Exhibits A-G). Indeed, Brown and Shalon U.S. Patent No. 5,807,522 (the Brown '522 patent, Bedilion Exhibit D), which issued from a patent application filed in June 1995 and was effectively published on December 29, 1995 as a result of the publication of a PCT counterpart application, shows that the Patent Office recognizes the patentable utility of the cDNA technology developed in the early to mid-1990s. As explained by Dr. Bedilion, among other things (Bedilion Declaration, ¶ 12):

> The Brown '522 patent further teaches that the "[m]icroarrays of immobilized nucleic acid sequences prepared in accordance with the invention" can be used in "numerous" genetic applications, including "monitoring of gene expression" applications (see Bedilion Tab D at col. 14, lines 36-42). The Brown '522 patent teaches (a) monitoring gene expression (i) in different tissue types, (ii) in different disease states, and (iii) in response to different drugs, and (b) that arrays disclosed therein may be used in toxicology studies (see Bedilion Tab D at col. 15, lines 13-18 and 52-58 and col. 18, lines 25-30).

Literature reviews published shortly after the filing of the Tang et al. '404 application describing the state of the art further confirm the claimed invention's utility. Rockett et al. confirm, for example, that the claimed invention is useful for differential expression analysis regardless of how expression is regulated:

> Despite the development of multiple technological advances which have recently brought the field of gene expression profiling to the forefront of molecular analysis, recognition of the importance of differential gene expression and characterization of differentially expressed genes has existed for many years.
>
> * * *
>
> Although differential expression technologies are applicable to a broad range of models, perhaps their most important advantage is that, in most cases, absolutely no prior knowledge of the specific genes which are up- or down-regulated is required.
>
> * * *
>
> Whereas it would be informative to know the identity and functionality of all genes up/down regulated by . . . toxicants, this would appear a longer term goal . . . . However, the current use of gene profiling yields a *pattern* of gene changes for a xenobiotic of unknown toxicity which may be matched to that of well characterized toxins, thus alerting the toxicologist to possible *in vivo* similarities between the unknown and the standard, thereby providing a platform for more extensive toxicological examination. (emphasis added)

Rockett et al., Differential gene expression in drug metabolism and toxicology: practicalities, problems and potential, 29 Xenobiotica No. 7, 655 (1999).

In another pre-March 8, 1999 article, Lashkari et al. state explicitly that sequences that are merely "predicted" to be expressed (predicted Open Reading Frames, or ORFs) – the claimed invention in fact is known to be expressed – have numerous uses:

> Efforts have been directed toward the amplification of each predicted ORF or any other region of the genome ranging from a few base pairs to several kilobase pairs. There are many uses for these amplicons– they can be cloned into standard vectors or specialized expression vectors, or can be cloned into other specialized vectors such as those used for two-hybrid analysis. The amplicons can also be used directly by, for example, arraying onto glass for expression analysis, for DNA binding assays, or for any direct DNA assay.

Lashkari et al., Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR, 94 Proc. Nat. Acad. Sci. 8945 (Aug. 1997) (emphasis added).

2. The Furness Declaration

In his Declaration, Mr. Furness explains the many reasons why a person skilled in the art who read the Tang et al. '404 application on March 8, 1999 would have understood that application to disclose the claimed polypeptide to be useful for a number of gene and protein expression monitoring applications, *e.g.*, in 2-D PAGE technologies, in connection with the development of drugs and the monitoring of the activity of such drugs. (Furness Declaration at, e.g., ¶¶ [11-15]). Much, but not all, of Mr. Furness' explanation concerns the use of the claimed polypeptide in the creation of protein expression maps using 2-D PAGE.

2-D PAGE technologies were developed during the 1980's. Since the early 1990's, 2-D PAGE has been used to create maps showing the differential expression of proteins in different cell types or in similar cell types in response to drugs and potential toxic agents. Each expression pattern reveals the state of a tissue or cell type in its given environment, *e.g.*, in the presence or absence of a drug. By comparing a map of cells treated with a potential drug candidate to a map of cells not treated with the candidate, for example, the potential toxicity of a drug can be assessed. Furness Declaration at ¶ [11].)

The claimed invention makes 2-D PAGE analysis a more powerful tool for toxicology and drug efficacy testing. A person of ordinary skill in the art can derive more information about the state or states or tissue or cell samples from 2-D PAGE analysis with the claimed invention than without it. As Mr. Furness explains:

> In view of the Tang et al. '404 application, the Wilkins article, and other related pre-March, 1999 publications, persons skilled in the art on March 8, 1999 clearly

> would have understood the Tang et al. '404 application to disclose the SEQ ID NO:12 polypeptide to be useful in 2-D PAGE analyses for the development of new drugs and monitoring the activities of drugs for such purposes as evaluating their efficacy and toxicity. . . . (Furness Declaration, ¶10)
>
> * * *
>
> Persons skilled in the art would appreciate that a 2-D PAGE map that utilized the SEQ ID NO:12 polypeptide sequence would be a more useful tool than a 2-D PAGE map that did not utilize this protein sequence in connection with conducting protein expression monitoring studies on proposed (or actual) drugs for treating cell proliferative disorders, immune system disorders, infections, and neuronal disorders for such purposes as evaluating their efficacy and toxicity. (Furness Declaration, ¶12)

Mr. Furness' observations are confirmed in the literature published before the filing of the patent application. Wilkins, for example, describes how 2-D gels are used to define proteins present in various tissues and measure their levels of expression, the data from which is in turn used in databases:

> For proteome projects, the aim of [computer-aided 2-D PAGE] analysis . . . is to catalogue all spots from the 2-D gel in a qualitative and if possible quantitative manner, so as to define the number of proteins present and their levels of expression. Reference gel images, constructed from one or more gels, for the basis of two-dimensional gel databases. (Wilkins, Tab C, p. 26).

B. The use of polynucleotides and polypetides expressed by humans as tools for toxicology testing, drug discovery, and the diagnosis of disease is now "well-established"

The technologies made possible by expression profiling using polynucleotides and polypeptides are now well-established. The technical literature recognizes not only the prevalence of these technologies, but also their unprecedented advantages in drug development, testing and safety assessment. These technologies include toxicology testing, as described by Bedilion and Furness in their Declarations.

Toxicology testing is now standard practice in the pharmaceutical industry. See, *e.g.*, John C. Rockett et al., *supra*:

> Knowledge of toxin-dependent regulation in target tissues is not solely an academic pursuit as much interest has been generated in the pharmaceutical industry to harness this technology in the early identification of toxic drug candidates, thereby shortening the developmental process and contributing substantially to the safety assessment of new drugs.

To the same effect are several other scientific publications, including Emile F. Nuwaysir et al., Microarrays and Toxicology: The Advent of Toxicogenomics, 24 Molecular Carcinogenesis 153 (1999); Sandra Steiner and N. Leigh Anderson, Expression profiling in toxicology -- potentials and limitations, 112-13 Toxicology Letters 467 (2000).

Nucleic acids useful for measuring the expression of whole classes of genes are routinely incorporated for use in toxicology testing. Nuwaysir et al. describes, for example, a Human ToxChip comprising 2089 human clones, which were selected

> for their well-documented involvement in basic cellular processes as well as their responses to different types of toxic insult. Included on this list are DNA replication and repair genes, apoptosis genes, and genes responsive to PAHs and dioxin-like compounds, peroxisome proliferators, estrogenic compounds, and oxidant stress. Some of the other categories of genes include transcription factors, oncogenes, tumor suppressor genes, cyclins, kinases, phosphatases, cell adhesion and motility genes, and homeobox genes. Also included in this group are 84 housekeeping genes, whose hybridization intensity is averaged and used for signal normalization of the other genes on the chip.

See also Table 1 of Nuwaysir et al. (listing additional classes of genes deemed to be of special interest in making a human toxicology microarray).

The more genes that are available for use in toxicology testing, the more powerful the technique. "Arrays are at their most powerful when they contain the entire genome of the species they are being used to study." John C. Rockett and David J. Dix, Application of DNA Arrays to Toxicology, 107 Environ. Health Perspec.681, No. 8 (1999). Control genes are carefully selected for their stability across a large set of array experiments in order to best study the effect of toxicological compounds. See attached email from the primary investigator on the Nuwaysir paper, Dr. Cynthia Afshari, to an Incyte employee, dated July 3, 2000, as well as the original message to which she was responding, indicating that even the expression of carefully selected control genes can be altered. Thus, there is no expressed gene which is irrelevant to screening for toxicological effects, and all expressed genes have a utility for toxicological screening.

In fact, the potential benefit to the public, in terms of lives saved and reduced health care costs, are enormous. Recent developments provide evidence that the benefits of this information are already beginning to manifest themselves. Examples include the following:

- In 1999, CV Therapeutics, an Incyte collaborator, was able to use Incyte gene expression technology, information about the structure of a known transporter gene, and chromosomal mapping location, to identify the key gene associated with Tangiers disease. This discovery took place over a matter of only a few weeks, due to the power of these new genomics technologies. The discovery received an award from the American Heart Association as one of the top 10 discoveries associated with heart disease research in 1999.

- In an April 9, 2000, article published by the Bloomberg news service, an Incyte customer stated that it had reduced the time associated with target discovery and validation from 36 months to 18 months, through use of Incyte's genomic information database. Other Incyte customers have privately reported similar experiences. The implications of this significant saving of time and expense for the number of drugs that may be developed and their cost are obvious.

- In a February 10, 2000, article in the *Wall Street Journal*, one Incyte customer stated that over 50 percent of the drug targets in its current pipeline were derived from the Incyte database. Other Incyte customers have privately reported similar experiences. By doubling the number of targets available to pharmaceutical researchers, Incyte genomic information has demonstrably accelerated the development of new drugs.

C. The similarity of the claimed polypeptide to another of undisputed utility demonstrates utility

Because there is a substantial likelihood that the claimed HCSRP is functionally related to the gp120 receptor, a polypeptide of undisputed utility, there is by implication a substantial likelihood that the claimed polypeptide and the polynucleotide that encodes it are similarly useful. Applicants need not show any more to demonstrate utility. *In re Brana*, 51 F.3d at 1567.

It is undisputed that the claimed polypeptide shown as SEQ ID NO:12 in the patent application and referred to as HCSRP-12 shares 84% sequence identity over 325 amino acid residues with the gp120 receptor (GENESEQ AAR32188, International Patent WO 93/01820). The gp120 receptor belongs to the C-lectin receptor family whose members are known to mediate cellular immunity in part through carbohydrate recognition on microorganisms. Members of this family have been shown to bind glycoproteins on the viral envelopes of human

immunodeficiency virus (HIV) and Ebola virus (enclosed references of Curtis et al., Turville et al., and Alvarez et al.). Indeed, HCSRP-12 shows homology to other members of the C-lectin receptor family. The attention of the Examiner is directed to Exhibit B from the response to the Office Action of January 15, 2003, which shows a recent BLAST analysis of SEQ ID NO:12. The top hits include human L-SIGN (g13383470 and human mDC-SIGN type I isoform (g1538306), which except for a few sequence insertions, share 99.7% identity with SEQ ID NO:12. Both L-SIGN and DC-SIGN are known to bind to HIV gp120 and Ebola virus glycoproteins (enclosed references of Turville et al., Bashirova et al., and Alvarez et al.). The alignment of HCSRP-12 with human L-SIGN and mDC-SIGN corroborates the original determination of the instant application that HCSRP-12 was correctly assigned to the class of receptors that bind to HIV envelope glycoprotein gp120.

The attention of the Examiner is directed to Exhibit C from the response to the Office Action of January 15, 2003, which shows that SEQ ID NO:12 contains a C-type lectin domain from about residue S211 to residue K317 as determined by recent HMMER, MOTIFS, and BLIMPS analyses. Exhibit D from the response to the Office Action of January 15, 2003 shows an alignment of SEQ ID NO:12 with the sequences of the gp120 receptor (GENESEQ AAR32188), a membrane-associated C-type lectin that binds human immunodeficiency virus envelope glycoprotein gp120 (g8572543), and L-SIGN (g13383470), performed using the program, MEGALIGN version 4.05 and the CLUSTAL V algorithm. This alignment shows the presence of conserved residues, particularly in the region of SEQ ID NO:12 corresponding to the lectin domain. In all of these proteins, a C-type lectin domain is believed to mediate carbohydrate recognition and binding to the HIV envelope glycoprotein, gp120.

The homology among these sequences is more than enough to demonstrate a reasonable probability that the utility of the gp120 receptor can be imputed to the claimed invention. It is well-known that the probability that two unrelated polypeptides share more than 40% sequence homology over 70 amino acid residues is exceedingly small (Brenner et. al., Proc. Natl. Acad. Sci. (1998) 95:6073-78). Given homology in excess of 40% over many more than 70 amino acid residues, the probability that the claimed polypeptide is related to the gp120 receptor is, accordingly, very high.

It was known in the art at the time the application was filed that C-lectin receptors such as the gp120 receptor could be useful for detection of virus, inhibition of viral infection, and for development of vaccines (enclosed references of Geijtenbeek et al. and International patent WO 93/01820). It was also known that infection with HIV is associated with an increased incidence of cancer, particularly with Kaposi's sarcoma and non-Hodgkin's lymphoma, and that gp120 plays a role in tumor metastasis (enclosed references of Scadden and Hodgson et al.). In addition, gp120 induces neuronal apoptosis and neuronal injury associated with neurodegenerative disorders caused by HIV infection (enclosed references of Kaul et al. and Corasaniti et al.). Because of the relationship between HCSRP-12 and the gp120 receptor and C-lectin receptor proteins as a class, persons skilled in the art at the time the application was filed would have considered HCSRP-12 to be an important and valuable tool for use in research on cell proliferative disorders, immune system disorders, infections, and neuronal disorders.

The Examiner must accept the Applicants' demonstration that the homology between the claimed invention and the gp120 receptor demonstrates utility by a reasonable probability unless the Examiner can demonstrate through evidence or sound scientific reasoning that a person of ordinary skill in the art would doubt utility. *See In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). The Examiner has not provided sufficient evidence or sound scientific reasoning to the contrary.

D. Objective evidence corroborates the utilities of the claimed invention

There is, in fact, no restriction on the kinds of evidence a Patent Examiner may consider in determining whether a "real-world" utility exists. Indeed, "real-world" evidence, such as evidence showing actual use or commercial success of the invention, can demonstrate conclusive proof of utility. *Raytheon v. Roper*, 220 USPQ2d 592 (Fed. Cir. 1983); *Nestle v. Eugene*, 55 F.2d 854, 856, 12 USPQ 335 (6th Cir. 1932). Indeed, proof that the invention is made, used or sold by any person or entity other than the patentee is conclusive proof of utility. *United States Steel Corp.* v. *Phillips Petroleum Co.*, 865 F.2d 1247, 1252, 9 USPQ2d 1461 (Fed. Cir. 1989).

Over the past several years, a vibrant market has developed for databases containing all expressed genes (along with the polypeptide translations of those genes), in particular genes having medical and pharmaceutical significance such as the instant sequence. (Note that while

the value in these databases is enhanced by their completeness, each sequence in them is independently valuable nonetheless.) The databases sold by Applicants' assignee, Incyte, include exactly the kinds of information made possible by the claimed invention, such as tissue and disease associations. Incyte sells its database containing the claimed sequence and millions of other sequences throughout the scientific community, including to pharmaceutical companies who use the information to develop new pharmaceuticals.

Both Incyte's customers and the scientific community have acknowledged that Incyte's databases have proven to be valuable in, for example, the identification and development of drug candidates. As Incyte adds information to its databases, including the information that can be generated only as a result of Incyte's discovery of the claimed polynucleotide and its use of that polynucleotide on cDNA microarrays, the databases become even more powerful tools. Thus the claimed invention adds more than incremental benefit to the drug discovery and development process.

III. The Patent Examiner's Rejections Are Without Merit

Rather than responding to the evidence demonstrating utility, the Examiner attempts to dismiss it altogether by arguing that the disclosed and well-established utilities for the claimed polynucleotide and polypeptide are not "specific and substantial" utilities. (Office Action at p. 5.) The Examiner is incorrect both as a matter of law and as a matter of fact.

A. The Precise Biological Role Or Function Of An Expressed Polynucleotide or Polypeptide Is Not Required To Demonstrate Utility

The Patent Examiner's primary rejection of the claimed invention is based on the ground that, without information as to the precise "biological role" of the claimed invention, the claimed invention's utility is not sufficiently specific. According to the Examiner, it is not enough that a person of ordinary skill in the art could use and, in fact, would want to use the claimed invention either by itself or in a microarray, 2-D gel or western blot to monitor the expression of genes for such applications as the evaluation of a drug's efficacy and toxicity. The Examiner would

require, in addition, that the Applicant provide a specific and substantial interpretation of the results generated in any given expression analysis.

It may be that specific and substantial interpretations and detailed information on biological function are necessary to satisfy the requirements for publication in some technical journals, but they are not necessary to satisfy the requirements for obtaining a United States patent. The relevant question is not, as the Examiner would have it, whether it is known how or why the invention works, *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999), but rather whether the invention provides an "identifiable benefit" in presently available form. *Juicy Whip Inc.* v. *Orange Bang Inc.*, 185 F.3d 1364, 1366 (Fed. Cir. 1999). If the benefit exists, and there is a substantial likelihood the invention provides the benefit, it is useful. There can be no doubt, particularly in view of the Bedilion Declaration (at, *e.g.*, ¶¶ 10 and 15, Bedilion) and the Furness Declaration (at, *e.g.*, ¶¶ 10-13), that the present invention meets this test.

The threshold for determining whether an invention produces an identifiable benefit is low. *Juicy Whip*, 185 F.3d at 1366. Only those utilities that are so nebulous that a person of ordinary skill in the art would not know how to achieve an identifiable benefit and, at least according to the PTO guidelines, so-called "throwaway" utilities that are not directed to a person of ordinary skill in the art at all, do not meet the statutory requirement of utility. Utility Examination Guidelines, 66 Fed. Reg. 1092 (Jan. 5, 2001).

Knowledge of the biological function or role of a biological molecule has never been required to show real-world benefit. In its most recent explanation of its own utility guidelines, the PTO acknowledged so much (66 F.R. at 1095):

> [T]he utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have specific and substantial utility because, *e.g.*, it hybridizes near a disease-associated gene or it has gene-regulating activity.

By implicitly requiring knowledge of biological function for any claimed nucleic acid, the Examiner has, contrary to law, elevated what is at most an evidentiary factor into an absolute requirement of utility. Rather than looking to the biological role or function of the claimed invention, the Examiner should have looked first to the benefits it is alleged to provide.

B. Membership in a Class of Useful Products Can Be Proof of Utility

Despite the uncontradicted evidence that the claimed polypeptide is related to the gp120 receptor, a member of the C-type lectin receptor family, whose members indisputably are useful, the Examiner refused to impute the utility of the gp120 receptor to HCSRP-12. In the Office Action of January 15, 2003, the Patent Examiner takes the position that unless Applicants can identify which particular biological function of the gp120 receptor is possessed by HCSRP-12, utility cannot be imputed.

In order to demonstrate utility by membership in a class, the law requires only that the class not contain a substantial number of useless members. So long as the class does not contain a substantial number of useless members, there is sufficient likelihood that the claimed invention will have utility, and a rejection under 35 U.S.C. § 101 is improper. That is true regardless of how the claimed invention ultimately is used and whether or not the members of the class possess one utility or many. *See Brenner* v. *Manson*, 383 U.S. 519, 532 (1966); *Application of Kirk*, 376 F.2d 936, 943 (CCPA 1967).

Membership in a "general" class is insufficient to demonstrate utility only if the class contains a sufficient number of useless members such that a person of ordinary skill in the art could not impute utility by a substantial likelihood. There would be, in that case, a substantial likelihood that the claimed invention is one of the useless members of the class. In the few cases in which class membership did not prove utility by substantial likelihood, the classes did in fact include predominately useless members. *E.g.*, *Brenner* (man-made steroids); *Kirk* (same); *Natta* (man-made polyethylene polymers).

The Examiner addresses HCSRP-12 as if the general class in which it is included is not the C-type lectin receptor family, but rather all polynucleotides or all polypeptides, including the vast majority of useless theoretical molecules not occurring in nature, and thus not pre-selected by nature to be useful. While these "general classes" may contain a substantial number of useless members, the C-type lectin receptor family does not. The C-type lectin receptor family is sufficiently specific to rule out any reasonable possibility that HCSRP-12 would not also be useful like the other members of the family.

Because the Examiner has not presented any evidence that the C-type lectin receptor class of proteins has any, let alone a substantial number, of useless members, the Examiner must conclude that there is a "substantial likelihood" that the HCSRP-12 encoded by the claimed polynucleotides is useful. It follows that SEQ ID NO:12 and SEQ ID NO:25 also are useful.

Even if the Examiner's "common utility" criterion were correct – and it is not – the gp120 receptor would meet it. It is undisputed that known members of the C-type lectin receptor family, including the gp120 receptor, function in cellular immunity and host defense against viral infections. A person of ordinary skill in the art need not know any more about how the claimed invention functions in cellular immunity and viral infections to use it, and the Examiner presents no evidence to the contrary. Instead, the Examiner makes the conclusory observation that a person of ordinary skill in the art would need to know whether, for example, any given gp120 receptor functions in cellular immunity or viral infections. The Examiner then goes on to assume that the only use for HCSRP-12 absent knowledge as to how HCSRP-12 actually works is further study of HCSRP-12 itself.

Not so. As demonstrated by Applicants, knowledge that HCSRP-12 is a C-type lectin related to the gp120 receptor is more than sufficient to make it useful for the diagnosis and treatment of cell proliferative disorders, immune system disorders, infections, and neuronal disorders. The Examiner must accept these facts to be true unless the Examiner can provide evidence or sound scientific reasoning to the contrary. But the Examiner has not done so.

C. Because the uses of polynucleotides encoding HCSRP in toxicology testing, drug discovery, and disease diagnosis are practical uses beyond mere study of the invention itself, the claimed invention has substantial utility.

The Examiner rejected the claims at issue on the ground that the use of an invention as tool for research is not a "substantial" use. Because the Examiner's rejection assumes a substantial overstatement of the law, and is incorrect in fact, it must be overturned.

There is no authority for the proposition that use as a tool for research is not a substantial utility. Indeed, the Patent Office has recognized that just because an invention is used in a research setting does not mean that it lacks utility (MPEP § 2107):

> Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are

> useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact "useful" in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified utility and inventions whose specific utility requires further research to identify or reasonably confirm.

The Patent Office's actual practice has been, at least until the present, consistent with that approach. It has routinely issued patents for inventions whose only use is to facilitate research, such as DNA ligases. These are acknowledged by the PTO's Training Materials themselves to be useful, as well as DNA sequences used, for example, as markers.

Only a limited subset of research uses are not "substantial" utilities: those in which the only known use for the claimed invention is to be an **object** of further study, thus merely inviting further research. This follows from *Brenner*, in which the U.S. Supreme Court held that a process for making a compound does not confer a substantial benefit where the only known use of the compound was to be the object of further research to determine its use. *Id.* at 535. Similarly, in *Kirk*, the Court held that a compound would not confer substantial benefit on the public merely because it might be used to synthesize some other, unknown compound that would confer substantial benefit. *Kirk*, 376 F.2d at 940, 945 ("What Applicants are really saying to those in the art is take these steroids, experiment, and find what use they do have as medicines."). Nowhere do those cases state or imply, however, that a material cannot be patentable if it has some other beneficial use in research.

Such beneficial uses beyond studying the claimed invention itself have been demonstrated, in particular those described in the Bedilion and Furness Declarations. The claimed invention is a tool, rather than an object, of research. The data generated in gene expression monitoring using the claimed invention as a tool is **not** used merely to study the claimed polynucleotide itself, but rather to study properties of tissues, cells, and potential drug candidates and toxins. Without the claimed invention, the information regarding the properties of tissues, cells, drug candidates and toxins is less complete.

Moreover, as discussed above in section II D., SEQ ID NO:12 shares homology with other members of the C-lectin family that bind to viral glycoproteins. Therefore, the skilled artisan would have considered HCSRP to be an important and valuable tool, in particular, for use in research on cell proliferative disorders, immune system disorders, infections, and neuronal

disorders. The claimed invention has numerous other uses as a research tool, each of which alone is a "substantial utility." These include uses such as diagnostic assays (e.g., pages 40-44), chromosomal markers (e.g., pages 44-45), ligand screening assays (e.g., page 33), and drug screening (page 45-46).

IV. By Requiring the Patent Applicant to Assert a Particular or Unique Utility, the Patent Examination Utility Guidelines and Training Materials Applied by the Patent Examiner Misstate the Law

There is an additional, independent reason to overturn the rejections: to the extent the rejections are based on Revised Interim Utility Examination Guidelines (64 FR 71427, December 21, 1999), the final Utility Examination Guidelines (66 FR 1092, January 5, 2001) and/or the Revised Interim Utility Guidelines Training Materials (USPTO Website www.uspto.gov, March 1, 2000), the Guidelines and Training Materials are themselves inconsistent with the law.

The Training Materials, which direct the Examiners regarding how to apply the Utility Guidelines, address the issue of specificity with reference to two kinds of asserted utilities: "specific" utilities which meet the statutory requirements, and "general" utilities which do not. The Training Materials define a "specific utility" as follows:

> A [specific utility] is *specific* to the subject matter claimed. This contrasts to *general* utility that would be applicable to the broad class of invention. For example, a claim to a polynucleotide whose use is disclosed simply as "gene probe" or "chromosome marker" would not be considered to be specific in the absence of a disclosure of a specific DNA target. Similarly, a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed.

The Training Materials distinguish between "specific" and "general" utilities by assessing whether the asserted utility is sufficiently "particular," *i.e.*, unique (Training Materials at p.52) as compared to the "broad class of invention." (In this regard, the Training Materials appear to parallel the view set forth in Stephen G. Kunin, Written Description Guidelines and Utility Guidelines, 82 J.P.T.O.S. 77, 97 (Feb. 2000) ("With regard to the issue of specific utility the question to ask is whether or not a utility set forth in the specification is *particular* to the claimed invention.")).

Such "unique" or "particular" utilities never have been required by the law. To meet the utility requirement, the invention need only be "practically useful," *Natta*, 480 F.2d 1 at 1397, and confer a "specific benefit" on the public. *Brenner*, 383 U.S. at 534. Thus, incredible "throw-away" utilities, such as trying to "patent a transgenic mouse by saying it makes great snake food," do not meet this standard. Karen Hall, Genomic Warfare, The American Lawyer 68 (June 2000) (quoting John Doll, Chief of the Biotech Section of USPTO).

This does not preclude, however, a general utility, contrary to the statement in the Training Materials where "specific utility" is defined (page 5). Practical real-world uses are not limited to uses that are unique to an invention. The law requires that the practical utility be "definite," not particular. *Montedison,* 664 F.2d at 375. Applicant is not aware of any court that has rejected an assertion of utility on the grounds that it is not "particular" or "unique" to the specific invention. Where courts have found utility to be too "general," it has been in those cases in which the asserted utility in the patent disclosure was not a practical use that conferred a specific benefit. That is, a person of ordinary skill in the art would have been left to guess as to how to benefit at all from the invention. In *Kirk*, for example, the CCPA held the assertion that a man-made steroid had "useful biological activity" was insufficient where there was no information in the specification as to how that biological activity could be practically used. *Kirk*, 376 F.2d at 941.

The fact that an invention can have a particular use does not provide a basis for requiring a particular use. *See Brana, supra* (disclosure describing a claimed antitumor compound as being homologous to an antitumor compound having activity against a "particular" type of cancer was determined to satisfy the specificity requirement). "Particularity" is not and never has been the *sine qua non* of utility; it is, at most, one of many factors to be considered.

As described *supra*, broad classes of inventions can satisfy the utility requirement so long as a person of ordinary skill in the art would understand how to achieve a practical benefit from knowledge of the class. Only classes that encompass a significant portion of nonuseful members would fail to meet the utility requirement. *Supra* § II.B.2 (*Montedison,* 664 F.2d at 374-75).

The Training Materials fail to distinguish between broad classes that convey information of practical utility and those that do not, lumping all of them into the latter, unpatentable category of "general" utilities. As a result, the Training Materials paint with too broad a brush. Rigorous-

ly applied, they would render unpatentable whole categories of inventions that heretofore have been considered to be patentable and that have indisputably benefitted the public, including the claimed invention. *See supra* § II.B. Thus the Training Materials cannot be applied consistently with the law.

Issue 2 - Whether claims 21-30 and 35-37 meet the enablement requirement of 35 U.S.C. § 112, first paragraph

To the extent the rejection of the claimed invention under 35 U.S.C. § 112, first paragraph, is based on the improper rejection for lack of utility under 35 U.S.C. § 101, it must be reversed.

The rejection set forth in the Office Action is based on the assertions discussed above, i.e., that the claimed invention lacks patentable utility. To the extent that the rejection under § 112, first paragraph, is based on the improper allegation of lack of patentable utility under § 101, it fails for the same reasons.

Issue 3 - Whether claims 21, 23, 26, 27, 28, 30, 35, and 37 meet the written description requirement of 35 U.S.C. § 112, first paragraph

A. The Specification provides an adequate written description of the claimed variants and fragments of SEQ ID NO:12 and SEQ ID NO:25

The requirements necessary to fulfill the written description requirement of 35 U.S.C. 112, first paragraph, are well established by case law.

> . . . the applicant must also convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession *of the invention*. The invention is, for purposes of the "written description" inquiry, *whatever is now claimed. Vas-Cath, Inc. v. Mahurkar*, 19 USPQ2d 1111, 1117 (Fed. Cir. 1991)

Attention is also drawn to the Patent and Trademark Office's own "Guidelines for Examination of Patent Applications Under the 35 U.S.C. Sec. 112, para. 1", published January 5, 2001, which provide that :

> An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics[42] which provide evidence that applicant was in possession of the claimed invention,[43] i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics

> when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.[44] What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail.[45] If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met.[46]

Thus, the written description standard is fulfilled by both what is specifically disclosed and what is conventional or well known to one skilled in the art.

SEQ ID NO:12 and SEQ ID NO:25 are specifically disclosed in the application (see, for example, pages 21-22). Variants of SEQ ID NO:12 and SEQ ID NO:25 are described, for example, at page 22, line 23 through page 23, line 4. Incyte clones in which the nucleic acids encoding the human HCSRP were first identified and libraries from which those clones were isolated are described, for example, at Table 1 of the Specification. Chemical and structural features of SEQ ID NO:12 are described, for example, in Table 2. Given SEQ ID NO:12 and SEQ ID NO:25, one of ordinary skill in the art would recognize naturally-occurring variants of SEQ ID NO:12 having 85% sequence identity to SEQ ID NO:12 and naturally-occurring variants of SEQ ID NO:25 having 85% sequence identity to SEQ ID NO:25. Accordingly, the Specification provides an adequate written description of the recited polynucleotide and polypeptide sequences.

The Office Action has further asserted that the claims are not supported by an adequate written description because "it cannot be established that a representative number of species have been disclosed to support the genus claim" (Office Action, page 10).

Such a position is believed to present a misapplication of the law.

1. The present claims specifically define the claimed genus through the recitation of chemical structure

Court cases in which "DNA claims" have been at issue (which are hence relevant to claims to proteins encoded by the DNA) commonly emphasize that the recitation of structural features or chemical or physical properties are important factors to consider in a written

description analysis of such claims. For example, in *Fiers v. Revel*, 25 USPQ2d 1601, 1606 (Fed. Cir. 1993), the court stated that:

> If a conception of a DNA requires a precise definition, such as by structure, formula, chemical name or physical properties, as we have held, then a description also requires that degree of specificity.

In a number of instances in which claims to DNA have been found invalid, the courts have noted that the claims attempted to define the claimed DNA in terms of functional characteristics without any reference to structural features. As set forth by the court in *University of California v. Eli Lilly and Co.*, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997):

> In claims to genetic material, however, a generic statement such as "vertebrate insulin cDNA" or "mammalian insulin cDNA," without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function.

Thus, the mere recitation of functional characteristics of a DNA, without the definition of structural features, has been a common basis by which courts have found invalid claims to DNA. For example, in *Lilly*, 43 USPQ2d at 1407, the court found invalid for violation of the written description requirement the following claim of U.S. Patent No. 4,652,525:

> 1. A recombinant plasmid replicable in procaryotic host containing within its nucleotide sequence a subsequence having the structure of the reverse transcript of an mRNA of a vertebrate, which mRNA encodes insulin.

In *Fiers*, 25 USPQ2d at 1603, the parties were in an interference involving the following count:

> A DNA which consists essentially of a DNA which codes for a human fibroblast interferon-beta polypeptide.

Party Revel in the *Fiers* case argued that its foreign priority application contained an adequate written description of the DNA of the count because that application mentioned a potential method for isolating the DNA. The Revel priority application, however, did not have a description of any particular DNA structure corresponding to the DNA of the count. The court therefore found that the Revel priority application lacked an adequate written description of the subject matter of the count.

Thus, in *Lilly* and *Fiers*, nucleic acids were defined on the basis of functional characteristics and were found not to comply with the written description requirement of 35 U.S.C. §112; *i.e.*, "an mRNA of a vertebrate, which mRNA encodes insulin" in *Lilly*, and "DNA which codes for a human fibroblast interferon-beta polypeptide" in *Fiers*. In contrast to the situation in *Lilly* and *Fiers*, the claims at issue in the present application define polynucleotides or polypeptides specifically in terms of chemical structure, rather than on functional characteristics. For example, the "variant language" of independent claims 21 and 30 recite chemical structure to define the claimed genus:

> 21. An isolated polypeptide selected from the group consisting of:...
> b) a polypeptide comprising a naturally occurring amino acid sequence at least 85% identical to the amino acid sequence of SEQ ID NO:12...
>
> 30. An isolated polynucleotide selected from the group consisting of:...
> b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 85% identical to the polynucleotide sequence of SEQ ID NO:25...

From the above it should be apparent that the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structures of SEQ ID NO:12 and SEQ ID NO:25. In the present case, there is no reliance merely on a description of functional characteristics of the polynucleotides or polypeptides recited by the claims. In fact, there is no recitation of functional characteristics. Moreover, if such functional recitations were included, it would add to the structural characterization of the recited polynucleotides or polypeptides. The polynucleotides or polypeptides defined in the claims of the present application recite structural features, and cases such as *Lilly* and *Fiers* stress that the recitation of structure is an important factor to consider in a written description analysis of claims of this type. By failing to base its written description inquiry "on whatever is now claimed," the Office Action failed to provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in *Lilly* and *Fiers*

2. The present claims do not define a genus which is "highly variant"

Furthermore, the claims at issue do not describe a genus which could be characterized as "highly variant." Available evidence illustrates that the claimed genus is of narrow scope.

In support of this assertion, the attention of the Examiner is directed to the enclosed reference by Brenner et al. ("Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships," Proc. Natl. Acad. Sci. USA (1998) 95:6073-6078). Through exhaustive analysis of a data set of proteins with known structural and functional relationships and with <90% overall sequence identity, Brenner et al. have determined that 30% identity is a reliable threshold for establishing evolutionary homology between two sequences aligned over at least 150 residues. (Brenner et al., pages 6073 and 6076.) Furthermore, local identity is particularly important in this case for assessing the significance of the alignments, as Brenner et al. further report that ≥40% identity over at least 70 residues is reliable in signifying homology between proteins. (Brenner et al., page 6076.)

The present application is directed, *inter alia*, to human cell surface receptor proteins related to the amino acid sequence of SEQ ID NO:12. In accordance with Brenner et al, naturally occurring molecules may exist which could be characterized as human cell surface receptor proteins and which have as little as 40% identity over at least 70 residues to SEQ ID NO:12. The "variant language" of the present claims recites, for example, polypeptides or polynucleotides encoding "a naturally-occurring amino acid sequence having at least 85% sequence identity to the sequence of SEQ ID NO:12" (note that SEQ ID NO:12 has 325 amino acid residues). This variation is far less than that of all potential human cell surface receptor proteins related to SEQ ID NO:12, i.e., those human cell surface receptor proteins having as little as 40% identity over at least 70 residues to SEQ ID NO:12.

3. The state of the art at the time of the present invention is further advanced than at the time of the *Lilly* and *Fiers* applications

In the *Lilly* case, claims of U.S. Patent No. 4,652,525 were found invalid for failing to comply with the written description requirement of 35 U.S.C. §112. The '525 patent claimed the benefit of priority of two applications, Application Serial No. 801,343 filed May 27, 1977, and Application Serial No. 805,023 filed June 9, 1977. In the *Fiers* case, party Revel claimed the

benefit of priority of an Israeli application filed on November 21, 1979. Thus, the written description inquiry in those case was based on the state of the art at essentially at the "dark ages" of recombinant DNA technology.

The present application has a priority date of March 8, 1999. Much has happened in the development of recombinant DNA technology in the 22 or more years from the time of filing of the applications involved in *Lilly* and *Fiers* and the present application. For example, the technique of polymerase chain reaction (PCR) was invented. Highly efficient cloning and DNA sequencing technology has been developed. Large databases of protein and nucleotide sequences have been compiled. Much of the raw material of the human and other genomes has been sequenced. With these remarkable advances one of skill in the art would recognize that, given the sequence information of SEQ ID NO:12 and SEQ ID NO:25, and the additional extensive detail provided by the subject application, the present inventors were in possession of the claimed polynucleotide and polypeptide variants at the time of filing of this application.

Issue 4 – Whether claims 21-30 and 35-37 meet the requirements of 35 U.S.C. § 112, second paragraph

Claims 21-30 and 35-37 were rejected under 35 U.S.C. § 112, second paragraph, based on the allegation that the recitation of the term "naturally occurring" is indefinite because "all of the sequences existing in nature have not been identified" (Final Office Action, page 15). Applicants contend that the term "naturally occurring" is a well-known term in the art which Applicants intended to be used in such context. As such, no further definition of the term is necessary (MPEP 2163 IIA3(a)):

> What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail. See *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94. If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met. See, e.g., *Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972) (stating "the description need not be in *ipsis verbis* [i.e., "in the same words"] to be sufficient").

One of ordinary skill in the art would recognize that "***a naturally occurring amino acid sequence***" as recited in claim 21 is one which occurs in nature. Through the process of natural selection, nature will have determined the appropriate amino acid sequences. Given the information provided by SEQ ID NO:12 and SEQ ID NO:25, one of skill in the art would be able to routinely obtain a polynucleotide encoding "a naturally occurring amino acid sequence at least 85% identical to the amino acid sequence of SEQ ID NO:12." For example, the identification of relevant polynucleotides could be performed by hybridization and/or PCR techniques that were well-known to those skilled in the art at the time the subject application was filed and/or described throughout the Specification of the instant application. See, e.g., page 29, lines 22-33; page 40, lines 13-30; and Example VI at page 51.

Contrary to the Examiner's assertions, the Specification, as originally filed, provides adequate support for claiming polypeptides comprising a naturally occurring amino acid sequences having 85% sequence identity to SEQ ID NO:12. For example:

> "HCSRP" refers to the amino acid sequences of substantially purified HCSRP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.
> (Specification, page 9, lines 7-9)

Clearly, this definition of HCSRP encompasses naturally occurring variants of SEQ ID NO:12 from different species. The Specification further describes the identification of variants of SEQ ID NO:25.

> In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HCSRP or closely related molecules may be used to identify nucleic acid sequences which encode HCSRP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding HCSRP, allelic variants, or related sequences.
>
> Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the HCSRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:14-26 or from genomic sequences including promoters, enhancers, and introns of the HCSRP gene. (Specification, at page 41, lines 13-23)

In another embodiment of the invention, nucleic acid sequences encoding HCSRP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) (Specification, at page 44, line 29 through page 45, line 1)

See also Example VI at page 51.

Naturally occurring or recombinant HCSRP is substantially purified by immunoaffinity chromatography using antibodies specific for HCSRP. An immunoaffinity column is constructed by covalently coupling anti-HCSRP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

(Specification, page 55, lines 27-31)

Therefore, one of skill in the art could readily recognize and isolate a polypeptide comprising a naturally occurring amino acid sequence at least 85% identical to the amino acid sequence of SEQ ID NO:12. For at least the reasons set forth above, withdrawal of the rejection under U.S.C. § 112, second paragraph is respectfully requested.

CONCLUSION

Applicants respectfully submit that rejections for lack of utility based, *inter alia*, on an allegation of "lack of specificity," as set forth in the Office Action and as justified in the Revised Interim and final Utility Guidelines and Training Materials, are not supported in the law. Neither are they scientifically correct, nor supported by any evidence or sound scientific reasoning. These rejections are alleged to be founded on facts in court cases such as *Brenner* and *Kirk*, yet those facts are clearly distinguishable from the facts of the instant application, and indeed most if not all nucleotide and protein sequence applications. Nevertheless, the PTO is attempting to mold the facts and holdings of these prior cases, "like a nose of wax,"[2] to target rejections of claims to polypeptide and polynucleotide sequences, as well as to claims to methods of detecting said polynucleotide sequences, where biological activity information has not been proven by laboratory experimentation, and they have done so by ignoring perfectly acceptable utilities fully disclosed in the specifications as well as well-established utilities known to those of skill in the art. As is disclosed in the specification, and even more clearly, as one of ordinary skill in the art would understand, the claimed invention has well-established, specific, substantial and credible utilities. The rejections are, therefore, improper and should be reversed.

Moreover, to the extent the above rejections were based on the Revised Interim and final Examination Guidelines and Training Materials, those portions of the Guidelines and Training Materials that form the basis for the rejections should be determined to be inconsistent with the law.

The written description rejections under 35 U.S.C. § 112, first paragraph, should be reversed based on at least the arguments presented above. The Examiner failed to base the written description inquiry "on whatever is now claimed." Consequently, the Examiner did not provide an appropriate analysis of the present claims and how they differ from those found not to satisfy the written description requirement in cases such as *Lilly* and *Fiers*. In particular, the claims of the subject application are fundamentally different from those found invalid in *Lilly* and *Fiers*. The subject matter of the present claims is defined in terms of the chemical structure

---

[2]"The concept of patentable subject matter under §101 is not 'like a nose of wax which may be turned and twisted in any direction * * *.' *White v. Dunbar*, 119 U.S. 47, 51." *(Parker v. Flook,* 198 USPQ 193 (US SupCt 1978))

of SEQ ID NO:12 and SEQ ID NO:25. The courts have stressed that structural features are important factors to consider in a written description analysis of claims to nucleic acids and proteins. In addition, the genus of polypeptides defined by the present claims is adequately described, as evidenced by Brenner et al. Furthermore, there have been remarkable advances in the state of the art since the *Lilly* and *Fiers* cases, and these advances were given no consideration whatsoever in the position set forth by the Examiner.

The rejection under 35 U.S.C. § 112, second paragraph, should also be reversed based on at least the arguments presented above.

In light of the above amendments and remarks, Applicants submit that the present application is fully in condition for allowance, and request that the Examiner withdraw the outstanding objections/rejections. Early notice to that effect is earnestly solicited.

If the Examiner contemplates other action, or if a telephone conference would expedite allowance of the claims, Applicants invite the Examiner to contact the undersigned at the number listed below.

Please charge Deposit Account No. **09-0108** in the amount of **$770.00** as set forth in the enclosed fee transmittal letter. If the USPTO determines that an additional fee is necessary, please charge any required fee to Deposit Account No. **09-0108.**

Respectfully submitted,

INCYTE CORPORATION

Date: 4 February 2004

James M. Verna, Ph.D.
Reg. No. 33,287
Direct Dial Telephone: (650) 845 -5415

Date: February 4, 2004

Jenny Buchbinder, Ph.D.
Reg. No. 48,588
Direct Dial Telephone: (650) 843-7212

Customer No.: 27904
3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 849-8886

Enclosures:

1. Rockett et al., Differential gene expression in drug metabolism and toxicology: practicalities, problems and potential, 29 Xenobiotica No. 7, 655 (1999)
2. Lashkari et al., Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR, 94 Proc. Nat. Acad. Sci. 8945 (Aug. 1997).
3. Emile F. Nuwaysir et al., Microarrays and Toxicology: The Advent of Toxicogenomics, 24 Molecular Carcinogenesis 153 (1999);
4. Sandra Steiner and N. Leigh Anderson, Expression profiling in toxicology -- potentials and limitations, 112-13 Toxicology Letters 467 (2000).
5. John C. Rockett and David J. Dix, Application of DNA Arrays to Toxicology, 107 Environ. Health Perspec.681, No. 8 (1999).
6. Email from the primary investigator on the Nuwaysir paper, Dr. Cynthia Afshari, to an Incyte employee, dated July 3, 2000, as well as the original message to which she was responding.
7. Brenner et al., Proc. Natl. Acad. Sci. 95:6073-78 (1998).
8. Turville et al., Blood 98:2482-2488 (2001).
9. Bashirova et al., J. Exper. Med. 193:671-678 (2001).
10. Alvarez et al., J. Virol. 76:6841-6844 (2002).
11. Geijtenbeek et al., Cell 100:587-597 (2000).
12. International patent WO 93/01820.

13. Curtis et al., Proc. Natl. Acad. Sci. 89:8356-8360 (1992).
14. Kaul and Lipton, Proc. Natl. Acad. Sci. 96:8212-8216 (1999).
15. Corasaniti et al., J. Neurochem. 79:1-8 (2001).
16. Scadden et al., Ann. Rev. Med. 54:285-303 (2003).
17. Hodgson et al., Brain Res. 781:244-251 (1998).
18. Declaration of Mr. Furness, under 37 C.F.R. § 1.132.
19. Declaration of Dr. Bedilion, under 37 C.F.R. § 1.132.
21. Exhibit A
20. Exhibit B
21. Exhibit C
22. Exhibit D

XENOBIOTICA , 1999, VOL. 29, NO. 7, 655–691

Differential gene expression in drug metabolism and toxicology: practicalities, problems and potential

JOHN C. ROCKETT†, DAVID J. ESDAILE‡
and G. GORDON GIBSON*

Molecular Toxicology Laboratory, School of Biological Sciences, University of Surrey, Guildford, Surrey, GU2 5XH, UK

Received January 8, 1999

1. An important feature of the work of many molecular biologists is identifying which genes are switched on and off in a cell under different environmental conditions or subsequent to xenobiotic challenge. Such information has many uses, including the deciphering of molecular pathways and facilitating the development of new experimental and diagnostic procedures. However, the student of gene hunting should be forgiven for perhaps becoming confused by the mountain of information available as there appears to be almost as many methods of discovering differentially expressed genes as there are research groups using the technique.

2. The aim of this review was to clarify the main methods of differential gene expression analysis and the mechanistic principles underlying them. Also included is a discussion on some of the practical aspects of using this technique. Emphasis is placed on the so-called 'open' systems, which require no prior knowledge of the genes contained within the study model. Whilst these will eventually be replaced by 'closed' systems in the study of human, mouse and other commonly studied laboratory animals, they will remain a powerful tool for those examining less fashionable models.

3. The use of suppression-PCR subtractive hybridization is exemplified in the identification of up- and down-regulated genes in rat liver following exposure to phenobarbital, a well-known inducer of the drug metabolizing enzymes.

4. Differential gene display provides a coherent platform for building libraries and microchip arrays of 'gene fingerprints' characteristic of known enzyme inducers and xenobiotic toxicants, which may be interrogated subsequently for the identification and characterization of xenobiotics of unknown biological properties.

Introduction

It is now apparent that the development of almost all cancers and many non-neoplastic diseases are accompanied by altered gene expression in the affected cells compared to their normal state (Hunter 1991, Wynford-Thomas 1991, Vogelstein and Kinzler 1993, Semenza 1994, Cassidy 1995, Kleinjan and Van Hegningen 1998). Such changes also occur in response to external stimuli such as pathogenic micro-organisms (Rohn *et al.* 1996, Singh *et al.* 1997, Griffin and Krishna 1998, Lunney 1998) and xenobiotics (Sewall *et al.* 1995, Dogra *et al.* 1998, Ramana and Kohli 1998), as well as during the development of undifferentiated cells (Hecht 1998, Rudin and Thompson 1998, Schneider-Maunoury *et al.* 1998). The potential medical and therapeutic benefits of understanding the molecular changes which occur in any given cell in progressing from the normal to the 'altered' state are enormous. Such profiling essentially provides a 'fingerprint' of each step of a

* Author for correspondence; e-mail: g.gibson@surrey.ac.uk
† Current Address: US Environmental Protection Agency, National Health and Environmental Effects, Research Laboratory, Reproductive Toxicology Division, Research Triangle Park, NC 27711, USA.
‡ Rhone-Poulenc Agrochemicals, Toxicology Department, Sophia-Antipolis, Nice, France.

Xenobiotica ISSN 0049–8254 print/ISSN 1366–5928 online © 1999 Taylor & Francis Ltd
http://www.tandf.co.uk/JNLS/xen.htm http://www.taylorandfrancis.com/JNLS/xen.htm

cell's development or response and should help in the elucidation of specific and sensitive biomarkers representing, for example, different types of cancer or previous exposure to certain classes of chemicals that are enzyme inducers.

In drug metabolism, many of the xenobiotic-metabolizing enzymes (including the well-characterized isoforms of cytochrome P450) are inducible by drugs and chemicals in man (Pelkonen *et al.* 1998), predominantly involving transcriptional activation of not only the cognate cytochrome P450 genes, but additional cellular proteins which may be crucial to the phenomenon of induction. Accordingly, the development of methodology to identify and assess the full complement of genes that are either up- or down-regulated by inducers are crucial in the development of knowledge to understand the precise molecular mechanisms of enzyme induction and how this relates to drug action. Similarly, in the field of chemical-induced toxicity, it is now becoming increasingly obvious that most adverse reactions to drugs and chemicals are the result of multiple gene regulation, some of which are causal and some of which are casually-related to the toxicological phenomenon *per se*. This observation has led to an upsurge in interest in gene-profiling technologies which differentiate between the control and toxin-treated gene pools in target tissues and is, therefore, of value in rationalizing the molecular mechanisms of xenobiotic-induced toxicity. Knowledge of toxin-dependent gene regulation in target tissues is not solely an academic pursuit as much interest has been generated in the pharmaceutical industry to harness this technology in the early identification of toxic drug candidates, thereby shortening the developmental process and contributing substantially to the safety assessment of new drugs. For example, if the gene profile in response to say a testicular toxin that has been well-characterized *in vivo* could be determined in the testis, then this profile would be representative of all new drug candidates which act via this specific molecular mechanism of toxicity, thereby providing a useful and coherent approach to the early detection of such toxicants. Whereas it would be informative to know the identity and functionality of all genes up/down regulated by such toxicants, this would appear a longer term goal, as the majority of human genes have not yet been sequenced, far less their functionality determined. However, the current use of gene profiling yields a *pattern* of gene changes for a xenobiotic of unknown toxicity which may be matched to that of well-characterized toxins, thus alerting the toxicologist to possible *in vivo* similarities between the unknown and the standard, thereby providing a platform for more extensive toxicological examination. Such approaches are beginning to gain momentum, in that several biotechnology companies are commercially producing 'gene chips' or 'gene arrays' that may be interrogated for toxicity assessment of xenobiotics. These chips consist of hundreds/thousands of genes, some of which are degenerate in the sense that not all of the genes are mechanistically-related to any one toxicological phenomenon. Whereas these chips are useful in broad-spectrum screening, they are maturing at a substantial rate, in that gene arrays are now becoming more specific, e.g. chips for the identification of changes in growth factor families that contribute to the aetiology and development of chemically-induced neoplasias.

Although documenting and explaining these genetic changes presents a formidable obstacle to understanding the different mechanisms of development and disease progression, the technology is now available to begin attempting this difficult challenge. Indeed, several 'differential expression analysis' methods have been developed which facilitate the identification of gene products that demonstrate

altered expression in cells of one population compared to another. These methods have been used to identify differential gene expression in many situations, including invading pathogenic microbes (Zhao *et al.* 1998), in cells responding to extracellular and intracellular microbial invasion (Duguid and Dinauer 1990, Ragno *et al.* 1997, Maldarelli *et al.* 1998), in chemically treated cells (Syed *et al.* 1997, Rockett *et al.* 1999), neoplastic cells (Liang *et al.* 1992, Chang and Terzaghi-Howe 1998), activated cells (Gurskaya *et al.* 1996, Wan *et al.* 1996), differentiated cells (Hara *et al.* 1991, Guimaraes *et al.* 1995a, b), and different cell types (Davis *et al.* 1984, Hedrick *et al.* 1984, Xhu *et al.* 1998). Although differential expression analysis technologies are applicable to a broad range of models, perhaps their most important advantage is that, in most cases, absolutely no prior knowledge of the specific genes which are up- or down-regulated is required.

The field of differential expression analysis is a large and complex one, with many techniques available to the potential user. These can be categorized into several methodological approaches, including:

(1) Differential screening,
(2) Subtractive hybridization (SH) (includes methods such as chemical cross-linking subtraction—CCLS, suppression-PCR subtractive hybridization—SSH, and representational difference analysis—RDA),
(3) Differential display (DD),
(4) Restriction endonuclease facilitated analysis (including serial analysis of gene expression—SAGE—and gene expression fingerprinting—GEF),
(5) Gene expression arrays, and
(6) Expressed sequence tag (EST) analysis.

The above approaches have been used successfully to isolate differentially expressed genes in different model systems. However, each method has its own subtle (and sometimes not so subtle) characteristics which incur various advantages and disadvantages. Accordingly, it is the purpose of this review to clarify the mechanistic principles underlying the main differential expression methods and to highlight some of the broader considerations and implications of this very powerful and increasingly popular technique. Specifically, we will concentrate on the so-called 'open' systems, namely those which do not require any knowledge of gene sequences and, therefore, are useful for isolating unknown genes. Two 'closed' systems (those utilising previously identified gene sequences), EST analysis and the use of DNA arrays, will also be considered briefly for completeness. Whilst emphasis will often be placed on suppression PCR subtractive hybridization (SSH, the approach employed in this laboratory), it is the aim of the authors to highlight, wherever possible, those areas of common interest to those who use, or intend to use, differential gene expression analysis.

Differential cDNA library screening (DS)

Despite the development of multiple technological advances which have recently brought the field of gene expression profiling to the forefront of molecular analysis, recognition of the importance of differential gene expression and characterization of differentially expressed genes has existed for many years. One of the original approaches used to identify such genes was described 20 years ago by St John and Davis (1979). These authors developed a method, termed 'differential plaque filter

hybridization', which was used to isolate galactose-inducible DNA sequences from yeast. The theory is simple: a genomic DNA library is prepared from normal, unstimulated cells of the test organism/tissue and multiple filter replicas are prepared. These replica blots are probed with radioactively (or otherwise) labelled complex cDNA probes prepared from the control and test cell mRNA populations. Those mRNAs which are differentially expressed in the treated cell population will show a positive signal only on the filter probed with cDNA from the treated cells. Furthermore, labelled cDNA from different test conditions can be used to probe multiple blots, thereby enabling the identification of mRNAs which are only up-regulated under certain conditions. For example, St John and Davis (1979) screened replica filters with acetate-, glucose- and galactose-derived probes in order to obtain genes induced specifically by galactose metabolism. Although groundbreaking in its time this method is now considered insensitive and time-consuming, as up to 2 months are required to complete the identification of genes which are differentially expressed in the test population. In addition, there is no convenient way to check that the procedure has worked until the whole process has been completed.

Subtractive Hybridization (SH)

The developing concept of differential gene expression and the success of early approaches such as that described by St John and Davis (1979) soon gave rise to a search for more convenient methods of analysis. One of the first to be developed was SH, numerous variations of which have since been reported (see below). In general, this approach involves hybridization of mRNA/cDNA from one population (tester) to excess mRNA/cDNA from another (driver), followed by separation of the unhybridized tester fraction (differentially expressed) from the hybridized common sequences. This step has been achieved physically, chemically and through the use of selective polymerase chain reaction (PCR) techniques.

Physical separation

Original subtractive hybridization technology involved the physical separation of hybridized common species from unique single stranded species. Several methods of achieving this have been described, including hydroxyapatite chromatography (Sargent and Dawid 1983), avidin-biotin technology (Duguid and Dinauer 1990) and oligodT-latex separation (Hara *et al.* 1991). In the first approach, common mRNA species are removed by cDNA (from test cells)-mRNA (from control cells) subtractive hybridization followed by hydroxyapatite chromatography, as hydroxyapatite specifically adsorbs the cDNA-mRNA hybrids. The unabsorbed cDNA is then used either for the construction of a cDNA library of differentially expressed genes (Sargent and Dawid 1983, Schneider *et al.* 1988) or directly as a probe to screen a preselected library (Zimmerman *et al.* 1980, Davis *et al.* 1984, Hedrick *et al.* 1984). A schematic diagram of the procedure is shown in figure 1.

Less rigorous physical separation procedures coupled with sensitivity enhancing PCR steps were later developed as a means to overcome some of the problems encountered with the hydroxyapatite procedure. For example, Daguid and Dinauer (1990) described a method of subtraction utilizing biotin-affinity systems as a means to remove hybridized common sequences. In this process, both the control and tester mRNA populations are first converted to cDNA and an adaptor ('oligovector',

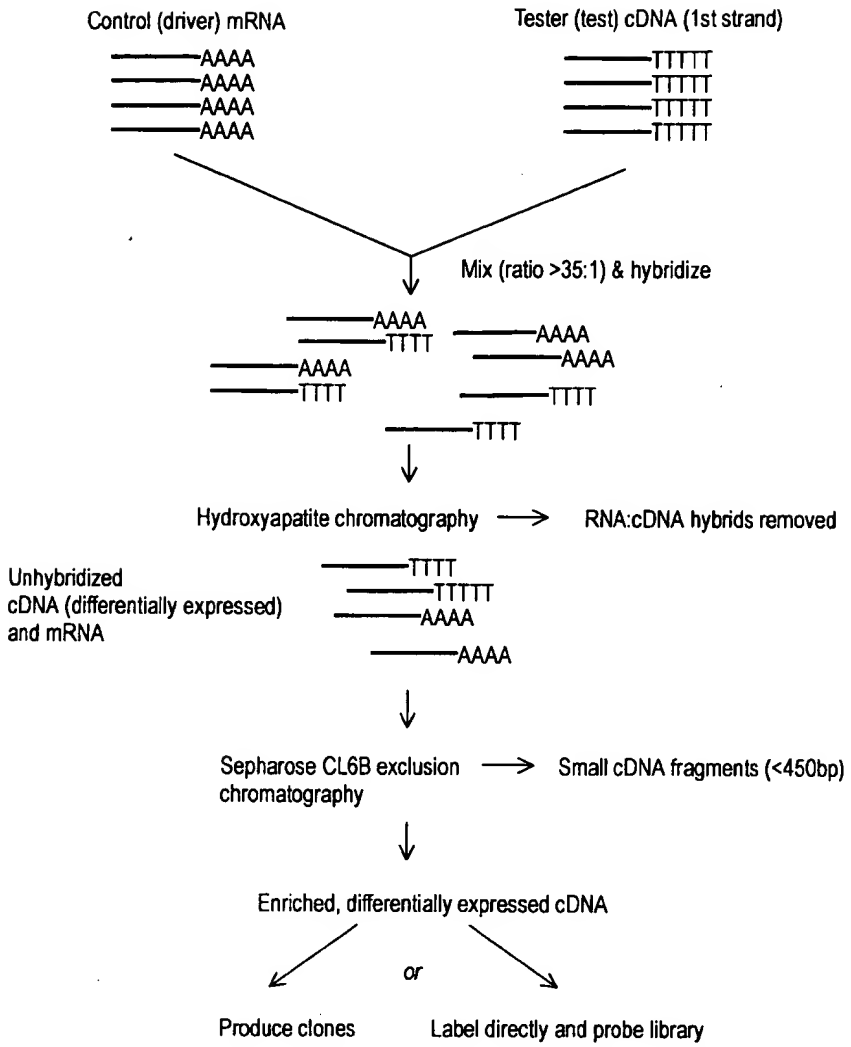


Figure 1. The hydroxyapatite method of subtractive hybridization. cDNA derived from the treated/altered (tester) population is mixed with a large excess of mRNA from the control (driver) population. Following hybridization, mRNA-cDNA hybrids are removed by hydroxyapatite chromatography. The only cDNAs which remain are those which are differentially expressed in the treated/altered population. In order to facilitate the recovery of full length clones, small cDNA fragments are removed by exclusion chromatography. The remaining cDNAs are then cloned into a vector for sequencing, or labelled and used directly to probe a library, as described by Sargent and Dawid (1983).

containing a restriction site) ligated to both sides. Both populations are then amplified by PCR, but the driver cDNA population is subsequently digested with the adaptor-containing restriction endonuclease. This serves to cleave the oligo-vector and reduce the amplification potential of the control population. The digested control population is then biotinylated and an excess mixed with tester cDNA. Following denaturation and hybridization, the mix is applied to a biocytin column (streptavidin may also be used) to remove the control population, including heteroduplexes formed by annealing of common sequences from the tester population. The procedure is repeated several times following the addition of fresh

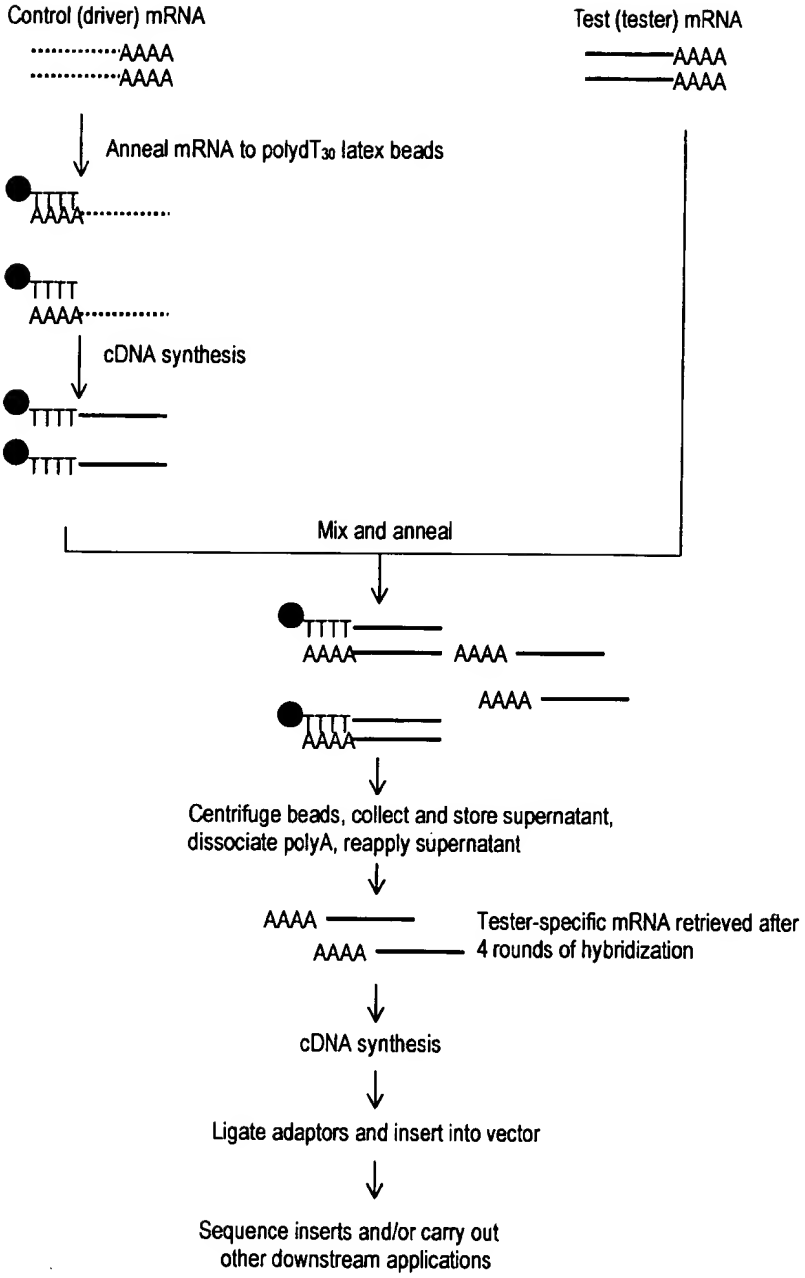


Figure 2. The use of oligodT_{30} latex to perform subtractive hybridization. mRNA extracted from the control (driver) population is converted to anchored cDNA using polydT oligonucleotides attached to latex beads. mRNA from the treated/altered (tester) population is repeatedly hybridized against an excess of the anchored driver cDNA. The final population of mRNA is tester specific and can be converted into cDNA for cloning and other downstream applications, as described by Hara *et al.* (1991).

control cDNA. In order to further enrich those species differentially expressed in the tester cDNA, the subtracted tester population is amplified by PCR following every second subtraction cycle. After six cycles of subtraction (three reamplification steps) the reaction mix is ligated into a vector for further analysis.

In a slightly different approach, Hara *et al.* (1991) utilized a method whereby oligo(dT_{30}) primers attached to a latex substrate are used to first capture mRNA extracted from the control population. Following 1st strand cDNA synthesis, the RNA strand of the heteroduplexes is removed by heat denaturation and centrifugation (the cDNA-oligotex-dT_{30} forms a pellet and the supernatant is removed). A quantity of tester mRNA is then repeatedly hybridized to the immobilized control (driver) cDNA (which is present in 20-fold excess). After several rounds of hybridization the only mRNA molecules left in the tester mRNA population are those which are not found in the driver cDNA-oligotex-dT_{30} population. These tester-specific mRNA species are then converted to cDNA and, following the addition of adaptor sequences, amplified by PCR. The PCR products are then ligated into a vector for further analysis using restriction sites incorporated into the PCR primers. A schematic illustration of this subtraction process is shown in figure 2.

However, all these methods utilising physical separation have been described as inefficient due to the requirement for large starting amounts of mRNA, significant loss of material during the separation process and a need for several rounds of hybridization. Hence, new methods of differential expression analysis have recently been designed to eliminate these problems.

Chemical Cross-Linking Subtraction (CCLS)

In this technique, originally described by Hampson *et al.* (1992), driver mRNA is mixed with tester cDNA (1st strand only) in a ratio of $> 20:1$. The common sequences form cDNA:mRNA hybrids, leaving the tester specific species as single stranded cDNA. Instead of physically separating these hybrids, they are inactivated chemically using 2,5 diaziridinyl-1,4-benzoquinone (DZQ). Labelled probes are then synthesized from the remaining single stranded cDNA species (unreacted mRNA species remaining from the driver are not converted into probe material due to specificity of Sequenase T7 DNA polymerase used to make the probe) and used to screen a cDNA library made from the tester cell population. A schematic diagram of the system is shown in figure 3.

It has been shown that the differentially expressed sequences can be enriched at least 300-fold with one round of subtraction (Hampson *et al.* 1992), and that the technique should allow isolation of cDNAs derived from transcripts that are present at less than 50 copies per cell. This equates to genes at the low end of intermediate abundance (see table 1). The main advantages of the CCLS approach are that it is rapid, technically simple and also produces fewer false positives than other differential expression analysis methods. However, like the physical separation protocols, a major drawback with CCLS is the large amount of starting material required (at least 10 μg RNA). Consequently, the technique has recently been refined so that a renewable source of RNA can be generated. The degenerate random oligonucleotide primed (DROP) adaptation (Hampson *et al.* 1996, Hampson and Hampson 1997) uses random hexanucleotide sequences to prime solid phase-synthesized cDNA. Since each primer includes a T7 polymerase promotor sequence

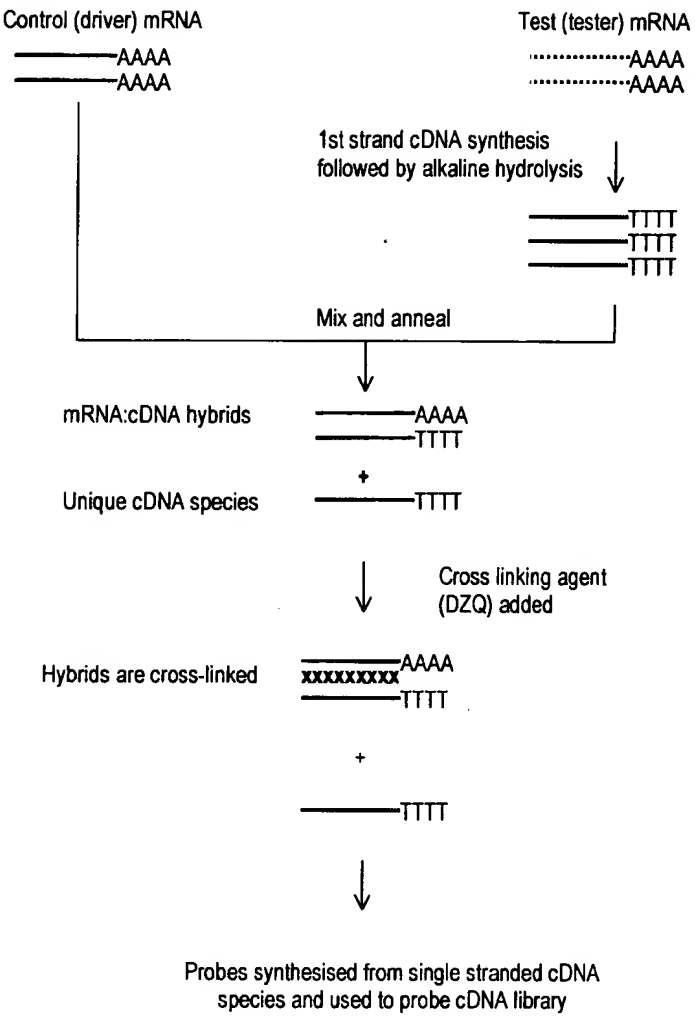


Figure 3. Chemical cross-linking subtraction. Excess driver mRNA is mixed with 1st strand tester cDNA. The common sequences form mRNA:cDNA hybrids which are cross linked with 2,5 diaziridinyl-1,4-benzoquinone (DZQ) and the remaining cDNA sequences are differentially expressed in the tester population. Probes are made from these sequences using Sequenase 2.0 DNA polymerase, which lacks reverse transcriptase activity and, therefore, does not react with the remaining mRNA molecules from the driver. The labelled probes are then used to screen a cDNA library for clones of differentially expressed sequences. Adapted from Walter *et al.* (1996), with permission.

Table 1. The abundance of mRNA species and classes in a typical mammalian cell.

mRNA class	Copies of each species/cell	No. of mRNA species in class	Mean % of each species in class	Mean mass (ng) of each species/μg total RNA
Abundant	12000	4	3.3	1.65
Intermediate	300	500	0.08	0.04
Rare	15	11000	0.004	0.002

Modified from Bertioli *et al.* (1995).

at the 5′ end, the final pool of random cDNA fragments is a PCR-renewable cDNA population which is representative of the expressed gene pool and can be used to synthesize sense RNA for use as driver material. Furthermore, if the final pool of random cDNA fragments is reamplified using biotinylated T7 primer and random hexamer, the product can be captured with streptavidin beads and the antisense strand eluted for use as tester. Since both target and driver can be generated from the same DROP product, subtraction can be performed in both directions (i.e. for up- and down-regulated species) between two different DROP products.

Representational Difference Analysis (RDA)

RDA of cDNA (Hubank and Schatz 1994) is an extension of the technique originally applied to genomic DNA as a means of identifying differences between two complex genomes (Lisitsyn *et al.* 1993). It is a process of subtraction and amplification involving subtractive hybridization of the tester in the presence of excess driver. Sequences in the tester that have homologues in the driver are rendered unamplifiable, whereas those genes expressed only in the tester retain the ability to be amplified by PCR. The procedure is shown schematically in figure 4.

In essence, the driver and tester mRNA populations are first converted to cDNA and amplified by PCR following the ligation of an adaptor. The adaptors are then removed from both populations and a new (different) adaptor ligated to the amplified tester population only. Driver and tester populations are next melted and hybridized together in a ratio of 100:1. Following hybridization, only tester:tester homohybrids have 5′ adaptors at each end of the DNA duplex and can, thus, be filled in at both 3′ ends. Hence, only these molecules are amplified exponentially during the subsequent PCR step. Although tester:driver heterohybrids are present, they only amplify in a linear fashion, since the strand derived from the driver has no adaptor to which the primer can bind. Driver:driver heterohybrids have no adaptors and, therefore, are not amplified. Single stranded molecules are digested with mung bean nuclease before a further PCR-enrichment of the tester:tester homohybrids. The adaptors on the amplified tester population are then replaced and the whole process repeated a further two or three times using an increasing excess of driver (Hubank and Shatz used a tester:driver ratio of 1:400, 1:80000 and 1:800000 for the second, third and fourth hybridizations, respectively). Different adaptors are ligated to the tester between successive rounds of hybridization and amplification to prevent the accumulation of PCR products that might interfere with subsequent amplifications. The final display is a series of differentially expressed gene products easily observable on an ethidium bromide gel.

The main advantages of RDA are that it offers a reproducible and sensitive approach to the analysis of differentially expressed genes. Hubank and Schatz (1994) reported that they were able to isolate genes that were differentially expressed in substantially less than 1% of the cells from which the tester is derived. Perhaps the main drawback is that multiple rounds of ligation, hybridization, amplifiation and digestion are required. The procedure is, therefore, lengthier than many other differential display approaches and provides more opportunity for operator-induced error to occur. Although the generation of false positives has been noted, this has been solved to some degree by O'Neill and Sinclair (1997) through the use of HPLC-purified adaptors. These are free of the truncated adaptors which appear to be a major source of the false positive bands. A very similar technique to RDA, termed linker capture subtraction (LCS) was described by Yang and Sytowski (1996).

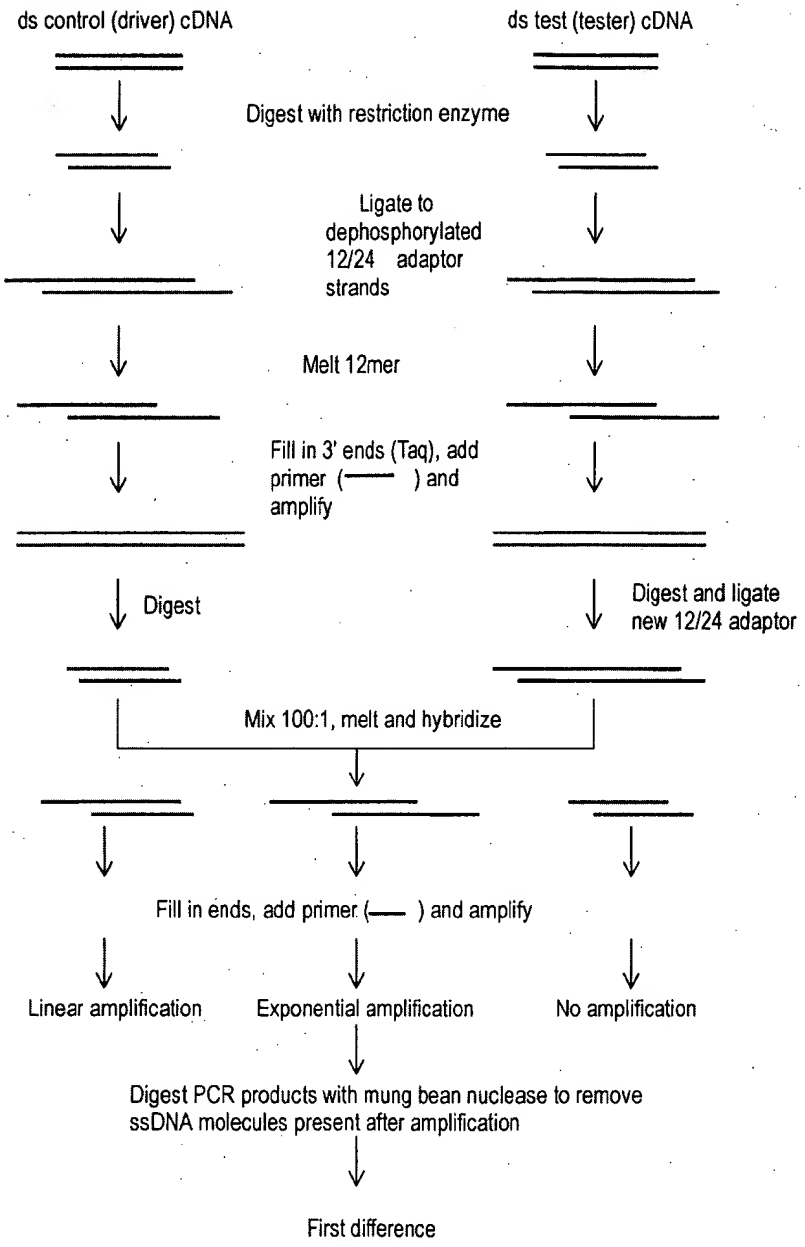


Figure 4. The representational difference analysis (RDA) technique. Driver and tester cDNA are digested with a 4-cutter restriction enzyme such as *Dpn*II. The 1st set of 12/24 adaptor strands (oligonucleotides) are ligated to each other and the digested cDNA products. The 12mer is subsequently melted away and the 3'ends filled in using Taq DNA polymerase. Each cDNA population is then amplified using PCR, following which the 1st set of adaptors is removed with *Dpn*II. A second set of 12/24 adaptor strands is then added to the amplified tester cDNA population, after which the tester is hybridized against a large excess of driver. The 12mer adaptors are melted and the 3´ends filled in as before. PCR is carried out with primers identical to the new 24mer adaptor. Thus, the only hybridization products which are exponentially amplified are those which are tester:tester combinations. Following PCR, ssDNA products are removed with mung bean nuclease, leaving the 'first difference product'. This is digested and a third set of 12/24 adaptors added before repeating the subtraction process from the hybridization stage. The process is repeated to the 3rd or 4th difference product, as described by Lisitsyn *et al.* (1993) and Hubank and Schatz (1994).

Suppression PCR Subtractive Hybridization (SSH)

The most recent adaptation of the SH approach to differential expression analysis was first described by Diatchenko *et al.* (1996) and Gurskaya *et al.* (1996). They reported that a 1000–5000 fold enrichment of rare cDNAs (equivalent to isolating mRNAs present at only a few copies per cell) can be obtained without the need for multiple hybridizations/subtractions. Instead of physical or chemical removal of the common sequences, a PCR-based suppression system is used (see figure 5).

In SSH, excess driver cDNA is added to two portions of the tester cDNA which have been ligated with different adaptors. A first round of hybridization serves to enrich differentially expressed genes and equalize rare and abundant messages. Equalization occurs since reannealing is more rapid for abundant molecules than for rarer molecules due to the second order kinetics of hybridization (James and Higgins 1985). The two primary hybridization mixes are then mixed together in the presence of excess driver and allowed to hybridize further. This step permits the annealing of single stranded complementary sequences which did not hybridize in the primary hybridization, and in doing so generates templates for PCR amplification. Although there are several possible combinations of the single stranded molecules present in the secondary hybridization mix, only one particular combination (differentially expressed in the tester cDNA composed of complimentary strands having different adaptors) can amplify exponentially.

Having obtained the final differential display, two options are available if cloning of cDNAs is desired. One is to transform the whole of the final PCR reaction into competent cells. Transformed colonies can then be isolated and their inserts characterized by sequencing, restriction analysis or PCR. Alternatively, the final PCR products can be resolved on a gel and the individual bands excised, reamplified and cloned. The first approach is technically simpler and less time consuming. However, ligation/transformation reactions are known to be biased towards the cloning of smaller molecules, and so the final population of clones will probably not contain a representative selection of the larger products. In addition, although equalization theoretically occurs, observations in this laboratory suggest that this is by no means perfectly accomplished. Consequently, some gene species are present in a higher number than others and this will be represented in the final population of clones. Thus, in order to obtain a substantial proportion of those gene species that actually demonstrate differential expression in the tester population, the number of clones that will have to be screened after this step may be substantial. The second approach is initially more time consuming and technically demanding. However, it would appear to offer better prospects for cloning larger and low abundance gel products. In addition, one can incorporate a screening step that differentiates different products of different sequences but of the same size (HA-staining, see later). In this way, a good idea of the final number of clones to be isolated and identified can be achieved.

An alternative (or even complementary) approach is to use the final differential display reaction to screen a cDNA library to isolate full length clones for further characterization, or a DNA array (see later) to quickly identify known genes. SSH has been used in this laboratory to begin characterization of the short-term gene expression profiles of enzyme-inducers such as phenobarbital (Rockett *et al.* 1997) and Wy-14,643 (Rockett *et al.* unpublished observations). The isolation of differentially expressed genes in this manner enables the construction of a fingerprint

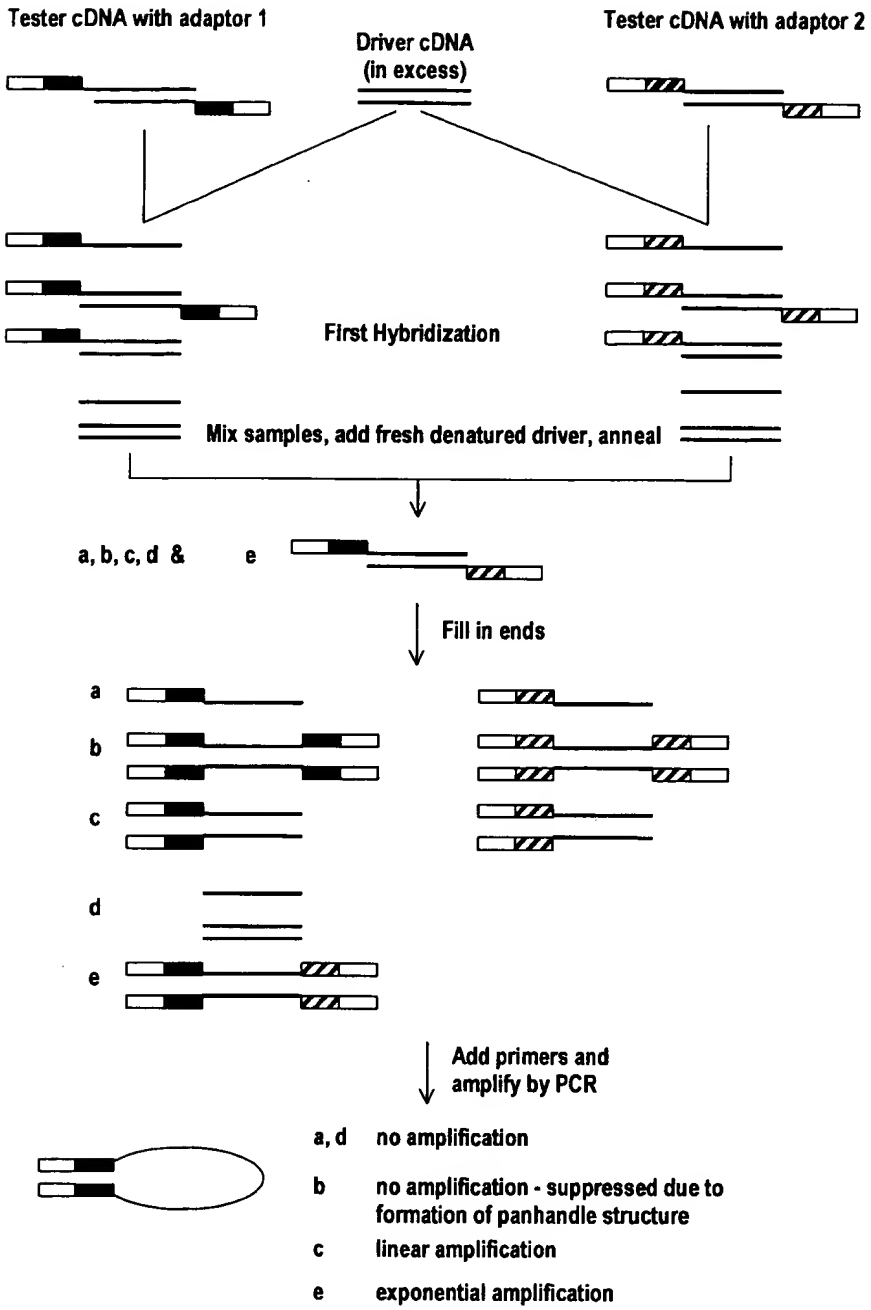


Figure 5. PCR-select cDNA subtraction. In the primary hybridization, an excess of driver cDNA is added to each tester cDNA population. The samples are heat denatured and allowed to hybridize for between 3 and 8 h. This serves two purposes: (1) to equalize rare and abundant molecules; and (2) to enrich for differentially expressed sequences—cDNAs that are not differentially expressed form type c molecules with the driver. In the secondary hybridization, the two primary hybridizations are mixed together without denaturing. Fresh denatured driver can also be added at this point to allow further enrichment of differentially expressed sequences. Type e molecules are formed in this secondary hybridization which are subsequently amplified using two rounds of PCR. The final products can be visualized on an agarose gel, labelled directly or cloned into a vector for downstream manipulation. As described by Diatchenko *et al.* (1996) and Gurskaya *et al.* (1996), with permission.

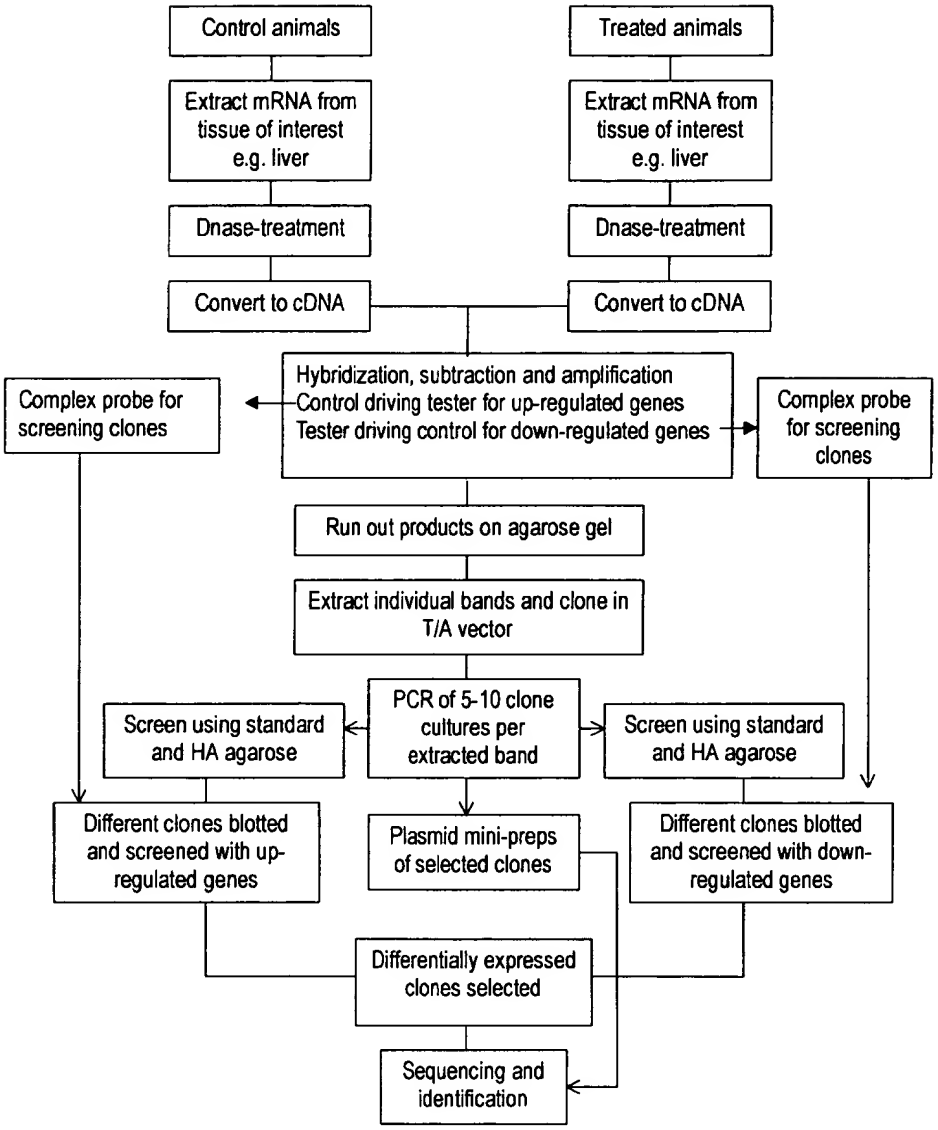


Figure 6. Flow diagram showing method used in this laboratory to isolate and identify clones of genes which are differentially expressed in rat liver following short term exposure to the enzyme inducers, phenobarbital and Wy-14,643.

of expressed genes which are unique to each compound and time/dose point. Such information could be useful in short-term characterization of the toxic potential of new compounds by comparing the gene-expression profiles they elicit with those produced by known inducers. Figure 6 shows a flow diagram of the method used to isolate, verify and clone differentially expressed genes, and figure 7 shows expression profiles obtained from a typical SSH experiment. Subsequent sub-cloning of the individual bands, sequencing and gene data base interrogation reveals many genes which are either up- or down-regulated by phenobarbital in the rat (tables 2 and 3).

One of the advantages in using the SSH approach is that no prior knowledge is required of which specific genes are up/down-regulated subsequent to xenobiotic

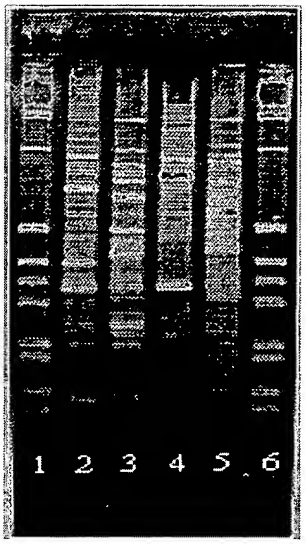


Figure 7. SSH display patterns obtained from rat liver following 3-day treatment with WY-14,643 or phenobarbital. mRNA extracted from control and treated livers was used to generate the differential displays using the PCR-Select cDNA subtraction kit (Clontech). Lane: 1—1kb ladder; 2—genes upregulated following Wy,14–643 treatment; 3—genes downregulated following Wy,14–643 treatment; 4—genes upregulated following phenobarbital treatment; 5—genes downregulated following phenobarbital treatment; 6—1kb ladder. Reproduced from Rockett *et al.* (1997), with permission.

exposure, and an almost complete complement of genes are obtained. For example, the peroxisome proliferator and non-genotoxic hepatocarcinogen Wy,14,643, up-regulates at least 28 genes and down-regulates at least 15 in the rat (a sensitive species) and produces 48 up- and 37 down-regulated genes in the guinea pig, a resistant species (Rockett, Swales, Esda and Gibson, unpublished observations). One of these genes, CD81, was up-regulated in the rat and down-regulated in the guinea pig following Wy-14,643 treatment. CD81 (alternatively named TAPA-1) is a widely expressed cell surface protein which is involved in a large number of cellular processes including adhesion, activation, proliferation and differentiation (Levy *et al.* 1998). Since all of these functions are altered to some extent in the phenomena of hepatomegaly and non-genotoxic hepatocarcinogenesis, it is intriguing, and probably mechanistically-relevant, that CD81 expression is differentially regulated in a resistant and susceptible species. However, the down-side of this approach is that the majority of genes can be sequenced and matched to database sequences, but the latter are predominantly expressed sequence tags or genes of completely unknown function, thus partially obscuring a realistic overall assessment of the critical genes of genuine biological interest. Notwithstanding the lack of complete funtional identification of altered gene expression, such gene profiling studies essentially provides a 'molecular fingerprint' in response to xenobiotic challenge, thereby serving as a mechanistically-relevant platform for further detailed investigations.

Differential Display (DD)

Originally described as 'RNA fingerprinting by arbitrarily primed PCR' (Liang and Pardee 1992) this method is now more commonly referred to as 'differential

Table 2. Genes up-regulated in rat liver following 3-day exposure to phenobarbital.

Band number (approximate size in bp)	Highest sequence similarity	FASTA-EMBL gene identification
5 (1300)	93.5%	CYP2B1
7 (1000)	95.1%	Preproalbumin Serum albumin mRNA
8 (950)	98.3%	NCI-CGAP-Pr1 *H. sapiens* (EST)
10 (850)	95.7%	CYP2B1
11 (800)	Clone 1 94.9%	CYP2B1
	Clone 2 75.3%	CYP2B2
12 (750)	93.8%	TRPM-2 mRNA Sulfated glycoprotein
15 (600)	92.9%	Preproalbumin Serum albumin mRNA
16 (55)	Clone 1 95.2%	CYP2B1
	Clone 2 93.6%	Haptoglobulin mRNA partial alpha
21 (350)	99.3%	18S, 5.8S & 28S rRNa

Bands 1–4, 6, 9, 13, 14, and 17–20 are shown to be false positives by dot blot anaylsis and, therefore, are not sequenced. Derived from Rockett *et al.* (1997). It should be noted that the above genes do not represent the complete spectrum of genes which are up-regulated in rat liver by phenobarbital, but simply represents the genes sequenced and identified to date.

Table 3. Genes down-regulated in rat liver following 3-day exposure to phenobarbital.

Band number (approximate size in bp)	Highest sequence similarity	FASTA-EMBL gene identification
1 (1500)	95.3%	3-oxoacyl-CoA thiolase
2 (1200)	92.3%	Hemopoxin mRNA
3 (1000)	91.7%	Alpha-2u-globulin mRNA
7 (700)	Clone 1 77.2%	*M.musculus* Cl inhibitor
	Clone 2 94.5%	Electron transfer flavoprotein
	Clone 3 91.0%	*M. musculus* Topoisomerase 1 (Topo 1)
8 (650)	Clone 1 86.9%	Soares 2NbMT *M. musculus* (EST)
	Clone 2 96.2%	Alpha-2u-globulin (s-type) mRNA
9 (600)	Clone 1 86.9%	Soares mouse NML *M. musculus* (EST)
	Clone 2 82.0%	Soares p3NMF 19.5 *M. musculus* (EST)
10 (550)	73.8%	Soares mouse NML *M. musculus* (EST)
11 (525)	95.7%	NCI-CGAP-Pr1 *H. sapiens* (EST)
12 (375)	100.0%	Ribosomal protein
13 (23)	Clone 1 97.2%	Soares mouse embryo NbME135 (EST)
	Clone 2 100.0%	Fibrinogen B-beta-chain
	Clone 3 100.0%	Apolipoprotein E gene
14 (170)	96.0%	Soares p3NMF19.5 *M. musculus* (EST)
15 (140)	97.3%	Stratagene mouse testis (EST)
Others: (300)	96.7%	*R. norvegicus* RASP 1 mRNA
(275)	93.1%	Soares mouse mammary gland (EST)

EST = Expressed sequence tag. Bands 4–6 were shown to be false positives by dot blot analysis and, therefore, were not sequenced. Derived from Rockett *et al.* (1997). It should be noted that the above genes do not represent the complete spectrum of genes which are down-regulated in rat liver by phenobarbital, but simiply represents the genes sequenced and identified to date.

display' (DD). In this method, all the mRNA species in the control and treated cell populations are amplified in separate reactions using reverse transcriptase-PCR (RT-PCR). The products are then run side-by-side on sequencing gels. Those bands which are present in one display only, or which are much more intense in one

display compared to the other, are differentially expressed and may be recovered for further characterization. One advantage of this system is the speed with which it can be carried out—2 days to obtain a display and as little as a week to make and identify clones.

Two commonly used variations are based on different methods of priming the reverse transcription step (figure 8). One is to use an oligo dT with a 2-base 'anchor' at the 3'-end, e.g. 5′ (dT_{11})CA 3′ (Liang and Pardee 1992). Alternatively, an arbitrary primer may be used for 1st strand cDNA synthesis (Welsh *et al.* 1992). This variant of RNA fingerprinting has also been called 'RAP' (RNA Arbitrarily Primed)-PCR. One advantage of this second approach is that PCR products may be derived from anywhere in the RNA, including open reading frames. In addition, it can be used for mRNAs that are not polyadenylated, such as many bacterial mRNAs (Wong and McClelland 1994). In both cases, following reverse transcription and denaturation, second strand cDNA synthesis is carried out with an arbitrary primer (*arbitrary* primers have a single base at each position, as compared to *random* primers, which contain a mixture of all four bases at each position). The resulting PCR, thus, produces a series of products which, depending on the system (primer length and composition, polymerase and gel system), usually includes 50–100 products per primer set (Band and Sager 1989). When a combination of different dT-anchors and arbitrary primers are used, almost all mRNA species from a cell can be amplified. When the cDNA products from two different populations are analysed side by side on a polyacrylamide gel, differences in expression can be identified and the appropriate bands recovered for cloning and further analysis.

Although DD is perhaps the most popular approach used today for identifying differentially expressed genes, it does suffer from several perceived disadvantages:

(1) It may have a strong bias towards high copy number mRNAs (Bertioli *et al.* 1995), although this has been disputed (Wan *et al.* 1996) and the isolation of very low abundance genes may be achieved in certain circumstances (Guimeraes *et al.* 1995a).
(2) The cDNAs obtained often only represent the extreme 3′ end of the mRNA (often the 3′-untranslated region), although this may not always be the case (Guimeraes *et al.* 1995a). Since the 3′ end is often not included in Genbank and shows variation between organisms, cDNAs identified by DD cannot always be matched with their genes, even if they have been identified.
(3) The pattern of differential expression seen on the display often cannot be reproduced on Northern blots, with false positives arising in up to 70% of cases (Sun *et al.* 1994). Some adaptations have been shown to reduce false positives, including the use of two reverse transcriptases (Sung and Denman 1997), comparison of uninduced and induced cells over a time course (Burn *et al.* 1994) and comparison of DDPCR-products from two uninduced and two induced lines (Sompayrac *et al.* 1995). The latter authors also reported that the use of cytoplasmic RNA rather then total RNA reduces false positives arising from nuclear RNA that is not transported to the cytoplasm.

Further details of the background, strengths and weaknesses of the DD technique can be obtained from a review by McClelland *et al.* (1996) and from articles by Liang *et al.* (1995) and Wan *et al.* (1996).

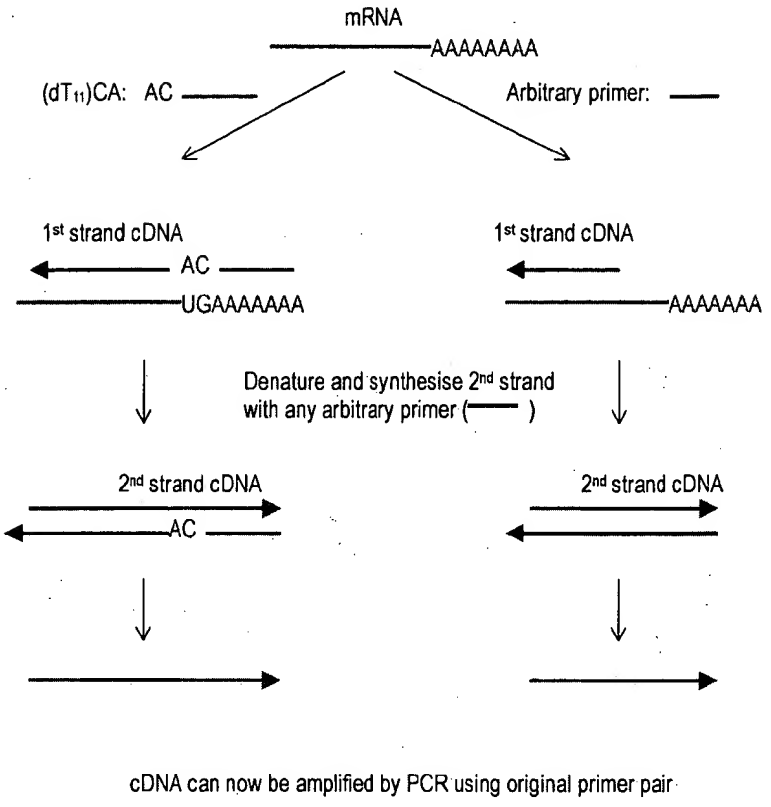


Figure 8. Two approaches to differential display (DD) analysis. 1st strand synthesis can be carried out either with a polydT$_{11}$NN primer (where N = G, C or A) or with an arbitrary primer. The use of different combinations of G, C and A to anchor the first strand polydT primer enables the priming of the majority of polyadenylated mRNAs. Arbitrary primers may hybridize at none, one or more places along the length of the mRNA, allowing 1st strand cDNA synthesis to occur at none, one or more points in the same gene. In both cases, 2nd strand synthesis is carried out with an arbitrary primer. Since these arbitrary primers for the 2nd strand may also hybridize to the 1st strand cDNA in a number of different places, several different 2nd strand products may be obtained from one binding point of the 1st strand primer. Following 2nd strand synthesis, the original set of primers is used to amplify the second strand products, with the result that numerous gene sequences are amplified.

Restriction endonuclease-facilitated analysis of gene expression

Serial Analysis of Gene Expression (SAGE)

A more recent development in the field of differential display is SAGE analysis (Velculescu *et al.* 1995). This method uses a different approach to those discussed so far and is based on two principles. Firstly, in more than 95% of cases, short nucleotide sequences ('tags') of only nine or 10 base pairs provide sufficient information to identify their gene of origin. Secondly, concatonation (linking together in a series) of these tags allows sequencing of multiple cDNAs within a single clone. Figure 9 shows a schematic representation of the SAGE process. In this procedure, double stranded cDNA from the test cells is synthesized with a biotinylated polydT primer. Following digestion with a commonly cutting (4bp recognition sequence) restriction enzyme ('anchoring enzyme'), the 3′ ends of the cDNA population are captured with streptavidin beads. The captured population is

split into two and different adaptors ligated to the 5′ ends of each group. Incorporated into the adaptors is a recognition sequence for a type IIS restriction enzyme—one which cuts DNA at a defined distance (< 20 bp) from its recognition sequence. Hence, following digestion of each captured cDNA population with the IIS enzyme, the adaptors plus a short piece of the captured cDNA are released. The two populations are then ligated and the products amplified. The amplified products are cleaved with the original anchoring enzyme, religated (concatomers are formed in the process) and cloned. The advantage of this system is that hundreds of gene tags can be identified by sequencing only a few clones. Furthermore, the number of times a given transcript is identified is a quantitative measurement of that gene's abundance in the original population, a feature which facilitates identification of differentially expressed genes in different cell populations.

Some disadvantages of SAGE analysis include the technical difficulty of the method, a large amount of accurate sequencing is required, biased towards abundant mRNAs, has not been validated in the pharmaco/toxicogenomic setting and has only been used to examine well known tissue differences to date.

Gene Expression Fingerprinting (GEF)

A different capture/restriction digest approach for isolating differentially expressed genes has been described by Ivanova and Belyavsky (1995). In this method, RNA is converted to cDNA using biotinylated oligo(dT) primers. The cDNA population is then digested with a specific endonuclease and captured with magnetic streptavidin microbeads to facilitate removal of the unwanted 5′ digestion products. The use of restricted 3′-ends alone serves to reduce the complexity of the cDNA fragment pool and helps to ensure that each RNA species is represented by not more than one restriction product. An adaptor is ligated to facilitate subsequent amplification of the captured population. PCR is carried out with one adaptor-specific and one biotinylated polydT primer. The reamplified population is recaptured and the non-biotinylated strands removed by alkaline dissociation. The non-biotinylated strand is then resynthesized using a different adaptor-specific primer in the presence of a radiolabelled dNTP. The labelled immobilized 3′ cDNA ends are next sequentially treated with a series of different restriction endonucleases and the products from each digestion analysed by PAGE. The result is a fingerprint composed of a number of ladders (equal to the number of sequential digests used). By comparing test versus control fingerprints, it is possible to identify differentially expressed products which can then be isolated from the gel and cloned. The advantages of this procedure are that it is very robust and reproducible, and the authors estimate that 80–93% of cDNA molecules are involved in the final fingerprint. The disadvantage is that polyacrylamide gels can rarely resolve more than 300–400 bands, which compares poorly to the 1000 or more which are estimated to be produced in an average experiment. The use of 2-D gels such as those described by Uitterlinden *et al.* (1989) and Hatada *et al.* (1991) may help to overcome this problem.

A similar method for displaying restriction endonuclease fragments was later described by Prashar and Weissman (1996). However, instead of sequential digestion of the immobolized 3'-terminal cDNA fragments, these authors simply compared the profiles of the control and treated populations without further manipulation.

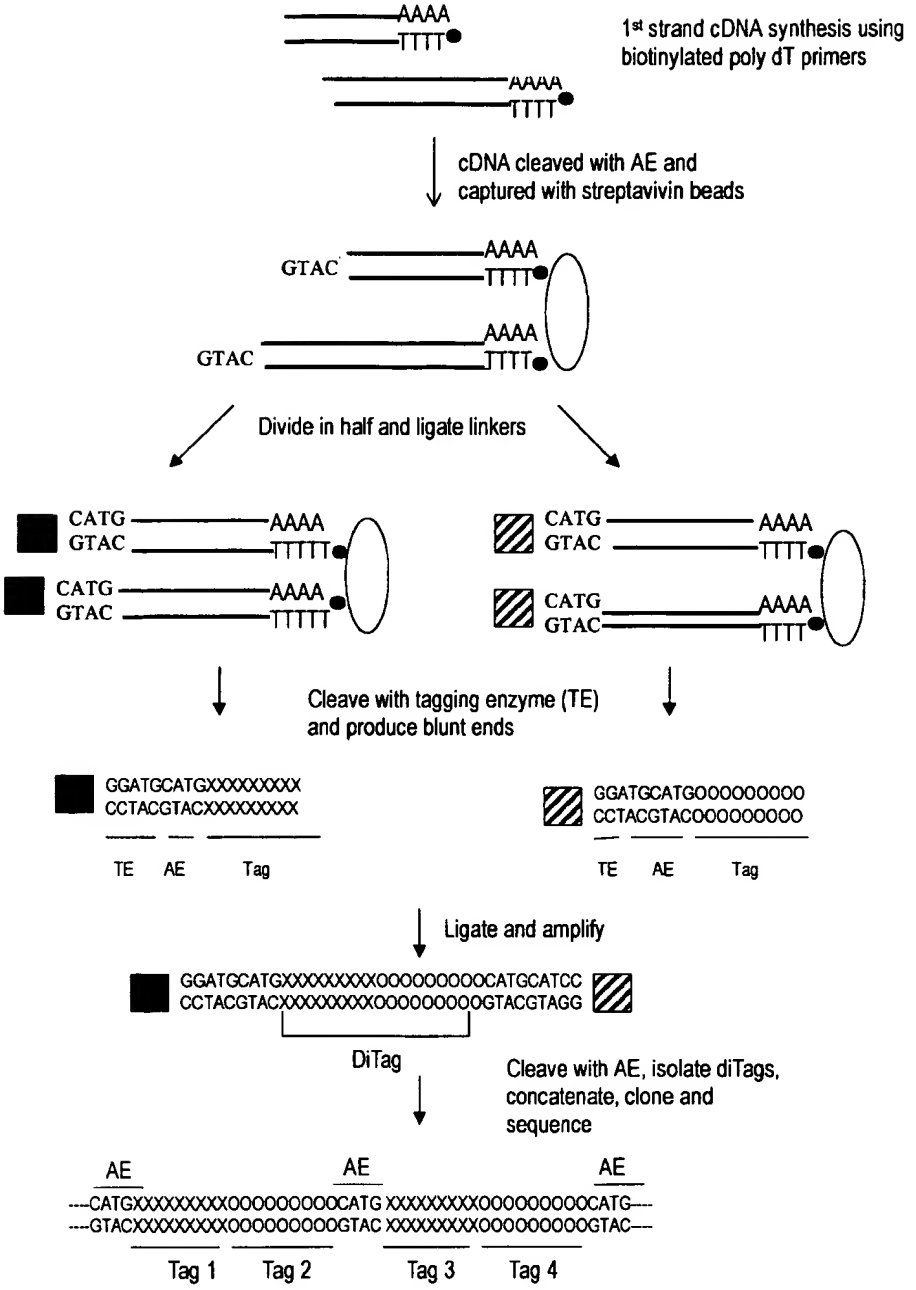


Figure 9. Serial analysis of gene expression (SAGE) analysis. cDNA is cleaved with an anchoring enzyme (AE) and the 3'ends captured using streptavidin beads. The cDNA pool is divided in half and each portion ligated to a different linker, each containing a type IIS restriction site (tagging enzyme, TE). Restriction with the type IIS enzyme releases the linker plus a short length of cDNA (XXXXX and OOOOO indicate nucleotides of different tags). The two pools of tags are then ligated and amplified using linker-specific primers. Following PCR, the products are cleaved with the AE and the ditags isolated from the linkers using PAGE. The ditags are then ligated (during which process, concatenization occurs) and cloned into a vector of choice for sequencing. After Velculescu *et al.* (1995), with permission.

DNA arrays

'Open' differential display systems are cumbersome in that it takes a great deal of time to extract and identify candidate genes and then confirm that they are indeed up- or down-regulated in the treated compared to the control tissue. Normally, the latter process is carried out using Northern blotting or RT-PCR. Even so, each of the aforementioned steps produce a bottleneck to the ultimate goal of rapid analysis of gene expression. These problems will likely be addressed by the development of so-called DNA arrays (e.g. Gress *et al.* 1992, Zhao *et al.* 1995, Schena *et al.* 1996), the introduction of which has signalled the next era in differential gene expression analysis. DNA arrays consist of a gridded membrane or glass 'chips' containing hundreds or thousands of DNA spots, each consisting of multiple copies of part of a known gene. The genes are often selected based on previously proven involvement in oncogenesis, cell cycling, DNA repair, development and other cellular processes. They are usually chosen to be as specific as possible for each gene and animal species. Human and mouse arrays are already commercially available and a few companies will construct a personalized array to order, for example Clontech Laboratories and Research Genetics Inc. The technique is rapid in that hundreds or even thousands of genes can be spotted on a single array, and that mRNA/cDNA from the test populations can be labelled and used directly as probe. When analysed with appropriate hardware and software, arrays offer a rapid and quantitative means to assess differences in gene expression between two cell populations. Of course, there can only be identification and quantitation of those genes which are in the array (hence the term 'closed' system). Therefore, one approach to elucidating the molecular mechanisms involved in a particular disease/development system may be to combine an open and closed system—a DNA array to directly identify and quantitate the expression of known genes in mRNA populations, and an open system such as SSH to isolate unknown genes which are differentially expressed.

One of the main advantages of DNA arrays is the huge number of gene fragments which can be put on a membrane—some companies have reported gridding up to 60000 spots on a single glass 'chip' (microscope slide). These high density chip-based micro-arrays will probably become available as mass-produced off-the-shelf items in the near future. This should facilitate the more rapid determination of differential expression in time and dose-response experiments. Aside from their high cost and the technical complexities involved in producing and probing DNA arrays, the main problem which remains, especially with the newer micro-array (gene-chip) technologies, is that results are often not wholly reproducible between arrays. However, this problem is being addressed and should be resolved within the next few years.

EST databases as a means to identify differentially expressed genes

Expressed sequence tags (ESTs) are partial sequences of clones obtained from cDNA libraries. Even though most ESTs have no formal identity (putative identification is the best to be hoped for), they have proven to be a rapid and efficient means of discovering new genes and can be used to generate profiles of gene-expression in specific cells. Since they were first described by Adams *et al.* (1991), there has been a huge explosion in EST production and it is estimated that there are now well over a million such sequences in the public domain, representing over half

of all human genes (Hillier *et al.* 1996). This large number of freely available sequences (both sequence information and clones are normally available royalty-free from the originators) has enabled the development of a new approach towards differential gene expression analysis as described by Vasmatzis *et al.* (1998). The approach is simple in theory: EST databases are first searched for genes that have a number of related EST sequences from the target tissue of choice, but none or few from non-target tissue libraries. Programmes to assist in the assembly of such sets of overlapping data may be developed in-house or obtained privately or from the internet. For example, the Institute for Genomic Research (TIGR, found at http://www.tigr.org) provides many software tools free of charge to the scientific community. Included amongst these is the TIGR assembler (Sutton *et al.* 1995), a tool for the assembly of large sets of overlapping data such as ESTs, bacterial artificial chromosomes (BAC)s, or small genomes. Candidate EST clones representing different genes are then analysed using RNA blot methods for size and tissue specificity and, if required, used as probes to isolate and identify the full length cDNA clone for further characterization. In practice however, the method is rather more involved, requiring bioinformatic and computer analysis coupled with confirmatory molecular studies. Vasmatzis *et al.* (1998) have described several problems in this fledgling approach, such as separating highly homologous sequences derived from different genes and an overemphasis of specificity for some EST sequences. However, since these problems will largely be addressed by the development of more suitable computer algorithms and an increased completeness of the EST database, it is likely that this approach to identifying differentially expressed genes may enjoy more patronage in the future.

Problems and potential of differential expression techniques

The holistic or single cell approach?

When working with *in vivo* models of differential expression, one of the first issues to consider must be the presence of multiple cell types in any given specimen. For example, a liver sample is likely to contain not only hepatocytes, but also (potentially) Ito cells, bile ductule cells, endothelial cells, various immune cells (e.g. lymphocytes, macrophages and Kupffer cells) and fibroblasts. Other tissues will each have their own distinctive cell populations. Also, in the case of neoplastic tissue, there are almost always normal, hyperplastic and/or dysplastic cells present in a sample. One must, therefore, be aware that genes obtained from a differential display experiment performed on an animal tissue model may not necessarily arise exclusively from the intended 'target' cells, e.g. hepatocytes/neoplastic cells. If appropriate, further analyses using immunohistochemistry, *in situ* hybridization or *in situ* RT-PCR should be used to confirm which cell types are expressing the gene(s) of interest. This problem is probably most acute for those studying the differential expression of genes in the development of different cell types, where there is a need to examine homologous cell populations. The problem is now being addressed at the National Cancer Institute (Bethesda, MD, USA) where new microdisection techniques have been employed to assist in their gene analysis programme, the Cancer Genome Anatomy Project (CGAP) (For more information see web site: http://www.ncbi.nlm.nih.gov/ncicgap/intro.html). There are also separation techniques available that utilise cell-specific antigens as a means to isolate target cells,

e.g. fluorescence activated cell sorting (FACS) (Dunbar *et al.* 1998, Kas-Deelen *et al.* 1998) and magnetic bead technology (Richard *et al.* 1998, Rogler *et al.* 1998).

However, those taking a holistic approach may consider this issue unimportant. There is an equally appropriate view that all those genes showing altered expression within a compromized tissue should be taken into consideration. After all, since all tissues are complex mixes of different, interacting cell types which intimately regulate each other's growth and development, it is clear that each cell type could in some way contribute (positively or negatively) towards the molecular mechanisms which lie behind responses to external stimuli or neoplastic growth. It is perhaps then more informative to carry out differential display experiments using *in vivo* as opposed to *in vitro* models, where uniform populations of identical cells probably represent a partial, skewed or even inaccurate picture of the molecular changes that occur.

The incidence and possible implications of inter-individual biological variation should be considered in any approach where whole animal models are being used. It is clear that individuals (humans and animals) respond in different ways to identical stimuli. One of the best characterized examples is the debrisoquine oxidation polymorphism, which is mediated by cytochrome CYP2D6 and determines the pharmacokinetics of many commonly prescribed drugs (Lennard 1993, Meyer and Zanger 1997). The reasons for such differences are varied and complex, but allelic variations, regulatory region polymorphisms and even physical and mental health can all contribute to observed differences in individual responses. Careful thought should, therefore, be given to the specific objectives of the study and to the possible value of pooling starting material (tissue/mRNA). The effect of this can be beneficial through the ironing out of exaggerated responses and unimportant minor fluctuations of (mechanistically) irrelevant genes in individual animals, thus providing a clearer overall picture of the general molecular mechanisms of the response. However, at the same time such minor variations may be of utmost importance in deciding the ability of individual animals to succumb to or resist the effects of a given chemical/disease.

How efficient are differential expression techniques at recovering a high percentage of differentially expressed genes?

A number of groups have produced experimental data suggesting that mammalian cells produce between 8000–15000 different mRNA species at any one time (Mechler and Rabbitts 1981, Hedrick *et al.* 1984, Bravo 1990), although figures as high as 20–30000 have also been quoted (Axel *et al.* 1976). Hedrick *et al.* (1984) provided evidence suggesting that the majority of these belong to the rare abundance class. A breakdown of this abundance distribution is shown in table 1.

When the results of differential display experiments have been compared with data obtained previously using other methods, it is apparent that not all differentially expressed mRNAs are represented in the final display. In particular, rare messages (which, importantly, often include regulatory proteins) are not easily recovered using differential display systems. This is a major shortcoming, as the majority of mRNA species exist at levels of less than 0.005% of the total population (table 1). Bertioli *et al.* (1995) examined the efficiency of DD templates (heterogeneous mRNA populations) for recovering rare messages and were unable to detect mRNA

species present at less than 1.2% of the total mRNA population—equivalent to an intermediate or abundant species. Interestingly, when simple model systems (single target only) were used instead of a heterogeneous mRNA population, the same primers could detect levels of target mRNA down to 10000× smaller. These results are probably best explained by competition for substrates from the many PCR products produced in a DD reaction.

The numbers of differentially expressed mRNAs reported in the literature using various model systems provides further evidence that many differentially expressed mRNAs are not recovered. For example, DeRisi *et al.* (1997) used DNA array technology to examine gene expression in yeast following exhaustion of sugar in the medium, and found that more than 1700 genes showed a change in expression of at least 2-fold. In light of such a finding, it would not be unreasonable to suggest that of the 8000–15 000 different mRNA species produced by any given mammalian cell, up to 1000 or more may show altered expression following chemical stimulation. Whilst this may be an extreme figure, it is known that at least 100 genes are activated/upregulated in Jurkat (T-) cells following IL-2 stimulation (Ullman *et al.* 1990). In addition, Wan *et al.* (1996) estimated that interferon-γ-stimulated HeLa cells differentially express up to 433 genes (assuming 24000 distinct mRNAs expressed by the cells). However, there have been few publications documenting anywhere near the recovery of these numbers. For example, in using DD to compare normal and regenerating mouse liver, Bauer *et al.* (1993) found only 70 of 38000 total bands to be different. Of these, 50% (35 genes) were shown to correspond to differentially expressed bands. Chen *et al.* (1996) reported 10 genes upregulated in female rat liver following ethinyl estradiol treatment. McKenzie and Drake (1997) identified 14 different gene products whose expression was altered by phorbol myristate acetate (PMA, a tumour promoter agent) stimulation of a human myelomonocytic cell line. Kilty and Vickers (1997) identified 10 different gene products whose expression was upregulated in the peripheral blood leukocytes of allergic disease sufferers. Linskens *et al.* (1995) found 23 genes differentially expressed between young and senescent fibroblasts. Techniques other than DD have also provided an apparent paucity of differentially expressed genes. Using SH for example, Cao *et al.* (1997) found 15 genes differentially expressed in colorectal cancer compared to normal mucosal epithelium. Fitzpatrick *et al.* (1995) isolated 17 genes upregulated in rat liver following treatment with the peroxisome proliferator, clofibrate; Philips *et al.* (1990) isolated 12 cDNA clones which were upregulated in highly metastatic mammary adenocarcinoma cell lines compared to poorly metastatic ones. Prashar and Weissman (1996) used 3′ restriction fragment analysis and identified approximately 40 genes showing altered expression within 4 h of activation of Jurkat T-cells. Groenink and Leegwater (1996) analysed 27 gene fragments isolated using SSH of delayed early response phase of liver regeneration and found only 12 to be upregulated.

In the laboratory, SSH was used to isolate up to 70 candidate genes which appear to show altered expression in guinea pig liver following short-term treatment with the peroxisome proliferator, WY-14,643 (Rockett, Swales, Esdaile and Gibson, unpublished observations). However, these findings have still to be confirmed by analysis of the extracted tissue mRNA for differential expression of these sequences.

Whilst the latest differential display technologies are purported to include design and experimental modifications to overcome this lack of efficiency (in both the total number of differentially expressed genes recovered and the percentage that are true

positives), it is still not clear if such adaptations are practically effective—proving efficiency by spiking with a known amount of limited numbers of artificial construct(s) is one thing, but isolating a high percentage of the rare messages already present in an mRNA population is another. Of course, some models will genuinely produce only a small number of differentially expressed genes. In addition, there are also technical problems that can reduce efficiency. For example, mRNAs may have an unusual primary structure that effectively prevents their amplification by PCR-based systems. In addition, it is known that under certain circumstances not all mRNAs have 3′polyA sites. For example, during *Xenopus* development, deadenylation is used as a means to stabilize RNAs (Voeltz and Steitz 1998), whilst preferential deadenylation may play a role in regulating Hsp70 (and perhaps, therefore, other stress protein) expression in *Drosophila* (Dellavalle *et al.* 1994). The presence of deadenylated mRNAs would clearly reduce the efficiency of systems utilizing a polydT reverse transcription step. The efficiency of any system also depends on the quality of the starting material. All differential display techniques use mRNA as their target material. However, it is difficult to isolate mRNA that is completely free of ribosomal RNA. Even if polydT primers are used to prime first strand cDNA synthesis, ribosomal RNA is often transcribed to some degree (Clontech PCR-Select cDNA Subtraction kit user manual). It has been shown, at least in the case of SSH, that a high rRNA:mRNA ratio can lead to inefficient subtractive hybridization (Clontech PCR-Select cDNA Subtraction kit user manual), and there is no reason to suppose that it will not do likewise in other SH approaches. Finally, those techniques that utilise a presubtraction amplification step (e.g. RDA) may present a skewed representation since some sequences amplify better than others.

Of course, probably the most important consideration is the temporal factor. It is clear that any given differential display experiment can only interrogate a cell at one point in time. It may well be that a high percentage of the genes showing altered expression at that time are obtained. However, given that disease processes and responses to environmental stimuli involve dynamic cascades of signalling, regulation, production and action, it is clear that all those genes which are switched on/off at different times will not be recovered and, therefore, vital information may well be missed. It is, therefore, imperative to obtain as much information about the model system beforehand as possible, from which a strategy can be derived for targeting specific time points or events that are of particular interest to the investigator. One way of getting round this problem of single time point analysis is to conduct the experiment over a suitable time course which, of course, adds substantially to the amount of work involved.

How sensitive are differential expression technologies?

There has been little published data that addresses the issue of how large the change in expression must be for it to permit isolation of the gene in question with the various differential expression technologies. Although the isolation of genes whose expression is changed as little as 1.5-fold has been reported using SSH (Groenink and Leegwater 1996), it appears that those demonstrating a change in excess of 5-fold are more likely to be picked up. Thus, there is a 'grey zone' in between where small changes could fade in and out of isolation between

experiments and animals. DD, on the other hand, is not subject to this grey zone since, unlike SH approaches, it does not amplify the difference in expression between two samples. Wan *et al.* (1996) reported that differences in expression of twofold or more are detectable using DD.

Resolution and visualization of differential expression products

It seems highly improbable with current technology that a gel system could be developed that is able to resolve all gene species showing altered expression in any given test system (be it SH- or DD-based). Polyacrylamide gel electrophoresis (PAGE) can resolve size differences down to 0.2% (Sambrook *et al.* 1989) and are used as standard in DD experiments. Even so, it is clear that a complex series of gene products such as those seen in a DD will contain unresolvable components. Thus, what appears to be one band in a gel may in fact turn out to be several. Indeed, it has been well documented (Mathieu-Daude *et al.* 1996, Smith *et al.* 1997) that a single band extracted from a DD often represents a composite of heterogeneous products, and the same has been found for SSH displays in this laboratory (Rockett *et al.* 1997). One possible solution was offered by Mathieu-Daude *et al.* (1996), who extracted and reamplified candidate bands from a DD display and used single strand conformation polymorphism (SSCP) analysis to confirm which components represented the truly differentially expressed product.

Many scientists often try to avoid the use of PAGE where possible because it is technically more demanding than agarose gel electrophoresis (AGE). Unfortunately, high resolution agarose gels such as Metaphor (FMC, Lichfield, UK) and AquaPor HR (National Diagnostics, Hessle, UK), whilst easier to prepare and manipulate than PAGE, can only separate DNA sequences which differ in size by around 1.5–2% (15–20 base pairs for a 1Kb fragment). Thus, SSH, RDA or other such products which differ in size by less than this amount are normally not resolvable. However, a simple technique does in fact exist for increasing the resolving power of AGE—the inclusion of HA-red (10-phenyl neutral red-PEG ligand) or HA-yellow (bisbenzamide-PEG ligand) (Hanse Analytik GmbH, Bremen, Germany) in a gel separates identical or closely sized products on base content. Specifically, HA-red and -yellow selectively bind to GC and AT DNA motifs, respectively (Wawer *et al.* 1995, Hanse Analytik 1997, personal communication). Since both HA-stains possess an overall positive charge, they migrate towards the cathode when an electric field is applied. This is in direct opposition to DNA, which is negatively charged and, therefore, migrates towards the anode. Thus, if two DNA clones are identical in size (as perceived on a standard high resolution agarose gel), but differ in AT/GC content, inclusion of a HA-dye in the gel will effectively retard the migration of one of the sequences compared to the other, effectively making it apparently larger and, thus, providing a means of differentiating between the two. The use of HA-red has been shown to resolve sequences with an AT variation of less than 1% (Wawer *et al.* 1995), whilst Hanse Analytik have reported that HA staining is so sensitive that in one case it was used to distinguish two 567bp sequences which differed by only a single point mutation (Hanse Analytik 1996, personal communication). Therefore, if one wishes to check whether all the clones produced from a specific band in a differential display experiment are derived from the same gene species, a small amount of reamplified or digested clone can be run on a standard high resolution gel, and a second aliquot

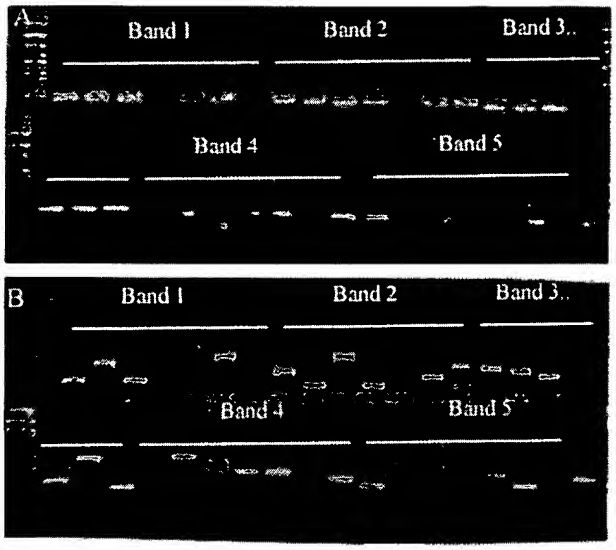


Figure 10. Discrimination of clones of identical/nearly identical size using HA-red. Bands of decreasing size (1–5) were extracted from the final display of a suppression subtractive hybridization experiment and cloned. Seven colonies were picked at random from each cloned band and their inserts amplified using PCR. The products were run on two gels, (A) a high resolution 2% agarose gel, and (B) a high resolution 2% agarose gel containing 1 U/ml HA-red. With few exceptions, all the clones from each band appear to be the same size (gel A). However, the presence of HA-red (gel B), which separates identically-sized DNA fragments based on the percentage of GC within the sequence, clearly indicates the presence of different gene species within each band. For example, even though all five re-amplified clones of band 1 appear to be the same size, at least four different gene species are represented.

in a similar gel containing one of the HA-stains. The standard gel should indicate any gross size differences, whilst the HA-stained gel should separate otherwise unresolvable species (on standard AGE) according to their base content. Geisinger *et al.* (1997) reported successful use of this approach for identifying DD-derived clones. Figure 10 shows such an experiment carried out in this laboratory on clones obtained from a band extracted from an SSH display.

An alternative approach is to carry out a 2-D analysis of the differential display products. In this approach, size-based separation is first carried out in a standard agarose gel. The gel slice containing the display is then extracted and incorporated in to a HA gel for resolution based on AT/GC content.

Of course, one should always consider the possibility of there being different gene species which are the same size and have the same GC/AT content. However, even these species are not unresolvable given some effort—again, one might use SSCP, or perhaps a denaturing gradient gel electrophoresis (DGGE) or temperature gradient field electrophoresis (TGGE) approach to resolve the contents of a band, either directly on the extracted band (Suzuki *et al.* 1991) or on the reamplified product.

The requirement of some differential display techniques to visualize large numbers of products (e.g. DD and GEF) can also present a problem in that, in terms of numbers, the resolution of PAGE rarely exceeds 300–400 bands. One approach to overcoming this might be to use 2-D gels such as those described by Uitterlinden *et al.* (1989) and Hatada *et al.* (1991).

Extraction of differentially expressed bands from a gel can be complex since, in some cases (e.g. DD, GEF), the results are visualized by autoradiographic means, such that precise overlay of the developed film on the gel must occur if the correct band is to be extracted for further analysis. Clearly, a misjudged extraction can account for many man-hours lost. This problem, and that of the use of radioisotopes, has been addressed by several groups. For example, Lohmann *et al.* (1995) demonstrated that silver staining can be used directly to visualize DD bands in horizontal PAGs. An *et al.* (1996) avoided the use of radioisotopes by transferring a small amount (20–30%) of the DNA from their DD to a nylon membrane, and visualizing the bands using chemiluminescent staining before going back to extract the remaining DNA from the gel. Chen and Peck (1996) went one step further and transferred the entire DD to a nylon membrane. The DNA bands were then visualized using a digoxigenin (DIG) system (DIG was attached to the polydT primers used in the differential display procedure). Differentially expressed bands were cut from the membrane and the DNA eluted by washing with PCR buffer prior to reamplification.

One of the advantages of using techniques such as SSH and RDA is that the final display can be run on an agarose gel and the bands visualized with simple ethidium bromide staining. Whilst this approach can provide acceptable results, overstaining with SYBR Green I or SYBR Gold nucleic acid stains (FMC) effectively enhances the intensity and sharpness of the bands. This greatly aids in their precise extraction and often reveals some faint products that may otherwise be overlooked. Whilst differential displays stained with SYBR Green I are better visualized using short wavelength UV (254 nm) rather than medium wavelength (306 nm), the shorter wavelength is much more DNA damaging. In practice, it takes only a few seconds to damage DNA extracted under 254 nm irradiation, effectively preventing reamplification and cloning. The best approach is to overstain with SYBR Green I and extract bands under a medium wavelength UV transillumination.

The possible use of 'microfingerprinting' to reduce complexity

Given the sheer number of gene products and the possible complexity of each band, an alternative approach to rapid characterization may be to use an enhanced analysis of a small section of a differential display—a 'sub-fingerprint' or 'micro-fingerprint'. In this case, one could concentrate on those bands which only appear in a particular chosen size region. Reducing the fingerprint in this way has at least two advantages. One is that it should be possible to use different gel types, concentrations and run times tailored exactly to that region. Currently, one might run products from 100–3000 +bp on the same gel, which leads to compromize in the gel system being used and consequently to suboptimal resolution, both in terms of size and numbers, and can lead to problems in the accurate excision of individual bands. Secondly, it may be possible to enhance resolution by using a 2-D analysis using a HA-stain, as described earlier. In summary, if a range of gene product sizes is carefully chosen to included certain 'relevant' genes, the 2-D system standardized, and appropriate gene analysis used, it may be possible to develop a method for the early and rapid identification of compounds which have similar or widely different cellular effects. If the prognosis for exposure to one or more other chemicals which display a similar profile is already known, then one could perhaps predict similar effects for any new compounds which show a similar micro-fingerprint.

An alternative approach to microfingerprinting is to examine altered expression in specific families of genes through careful selection of PCR primers and/or post-reaction analysis. Stress genes, growth factors and/or their receptors, cell cycling genes, cytochromes P450 and regulatory proteins might be considered as candidates for analysis in this way. Indeed, some off-the-shelf DNA arrays (e.g. Clontech's Atlas cDNA Expression Array series) already anticipated this to some degree by grouping together genes involved in different responses e.g. apoptosis, stress, DNA-damage response etc.

Screening

False positives

The generation of false positives has been discussed at length amongst the differential display community (Liang *et al.* 1993, 1995, Nishio *et al.* 1994, Sun *et al.* 1994, Sompayrac *et al.* 1995). The reason for false positives varies with the technique being used. For instance, in RDA, the use of adaptors which have not been HPLC purified can lead to the production of false positives through illegitimate ligation events (O'Neill and Sinclair 1997), whilst in DD they can arise through PCR artifacts and illegitemate transcription of rRNA. In SH, false positives appear to be derived largely from abundant gene species, although some may arise from cDNA/mRNA species which do not undergo hybridization for technical reasons.

A quick screening of putative differentially expressed clones can be carried out using a simple dot blot approach, in which labelled first strand probes synthesized from tester and driver mRNA are hybridized to an array of said clones (Hedrick *et al.* 1984, Sakaguchi *et al.* 1986). Differentially expressed clones will hybridize to tester probe, but not driver. The disadvantage of this approach is that rare species may not generate detectable hybridization signals. One option for those using SSH is to screen the clones using a labelled probe generated from the subtracted cDNA from which it was derived, and with a probe made from the reverse subtraction reaction (ClonTechniques 1997a). Since the SSH method enriches rare sequences, it should be possible to confirm the presence of clones representing low abundance genes. Despite this quick screening step, there is still the need to go back to the original mRNA and confirm the altered expression using a more quantitative approach. Although this may be achieved using Northern blots, the sensitivity is poor by today's high standards and one must rely on PCR methods for accurate and sensitive determinations (see below).

Sequence analysis

The majority of differential display procedures produce final products which are between 100 and 1000bp in size. However, this may considerably reduce the size of the sequence for analysis of the DNA databases. This in turn leads to a reduced confidence in the result—several families of genes have members whose DNA sequences are almost identical except in a few key stretches, e.g. the cytochrome P450 gene superfamily (Nelson *et al.* 1996). Thus, does the clone identified as being almost identical to gene X_0 really come from that gene, or its brother gene X_1 or its as yet undiscovered sister X_2? For example, using SSH, part of a gene was isolated,

which was up-regulated in the liver of rats exposed to Wy-14,643 and was identified by a FASTA search as being transferrin (data not shown). However, transferrin is known to be downregulated by hypolipidemic peroxisome proliferators such as Wy-14,643 (Hertz *et al.* 1996), and this was confirmed with subsequent RT-PCR analysis. This suggests that the gene sequence isolated may belong to a gene which is closely related to transferrin, but is regulated by a different mechanism.

A further problem associated with SH technology is redundancy. In most cases before SH is carried out, the cDNA population must first be simplified by restriction digestion. This is important for at least two reasons:

(1) To reduce complexity—long cDNA fragments may form complex networks which prevent the formation of appropriate hybrids, especially at the high concentrations required for efficient hybridization.
(2) Cutting the cDNAs into small fragments provides better representation of individual genes. This is because genes derived from related but distinct members of gene families often have similar coding sequences that may cross-hybridize and be eliminated during the subtraction procedure (Ko 1990). Furthermore, different fragments from the same cDNA may differ considerably in terms of hybridization and amplification and, thus, may not efficiently do one or the other (Wang and Brown 1991). Thus, some fragments from differentially expressed cDNAs may be eliminated during subtractive hybridization procedures. However, other fragments may be enriched and isolated. As a consequence of this, some genes will be cut one or more times, giving rise to two or more fragments of different sizes. If those same genes are differentially expressed, then two or more of the different size fragments may come through as separate bands on the final differential display, increasing the observed redundancy and increasing the number of redundant sequencing reactions.

Sequence comparisons also throw up another important point—at what degree of sequence similarity does one accept a result. Is 90% identitiy between a gene derived from your model species and another acceptably close? Is 95% between your sequence and one from the same species also acceptable? This problem is particularly relevant when the forward and reverse sequence comparisons give similar sequences with completely different gene species! An arbitrary decision seems to be to allocate genes that are definite (95% and above similarity) and then group those between 60 and 95% as being related or possible homologues.

Quantitative analysis

At some point, one must give consideration to the quantitative analysis of the candidate genes, either as a means of confirming that they are truly differentially expressed, or in order to establish just what the differences are. Northern blot analysis is a popular approach as it is relatively easy and quick to perform. However, the major drawback with Northern blots is that they are often not sensitive enough to detect rare sequences. Since the majority of messages expressed in a cell are of low abundance (see table 1), this is a major problem. Consequently, RT-PCR may be the method of choice for confirming differential expression. Although the procedure is somewhat more complex than Northern analysis, requiring synthesis of primers and optimization of reaction conditions for each gene species, it is now possible to set up high throughput PCR systems using mulitchannel pipettes, 96 +-well plates and

appropriate thermal cycling technology. Whilst quantitative analysis is more desirable, being more accurate and without reliance on an internal standard, the money and time needed to develop a competitor molecule is often excessive, especially when one might be examining tens or even hundreds of gene species. The use of semi-quantitative analysis is simpler, although still relatively involved. One must first of all choose an internal standard that does not change in the test cells compared to the controls. Numerous reference genes have been tried in the past, for example interferon-gamma (IFN-γ, Frye *et al.* 1989), β-actin (Heuval *et al.* 1994), glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Wong *et al.* 1994), dihydrofolate reductase (DHFR, Mohler and Butler 1991), β-2-microglobulin (β-2-m, Murphy *et al.* 1990), hypoxanthine phosphoribosyl transferase (HPRT, Foss *et al.* 1998) and a number of others (ClonTechniques 1997b). Ideally, an internal standard should not change its level of expression in the cell regardless of cell age, stage in the cell cycle or through the effects of external stimuli. However, it has been shown on numerous occasions that the levels of most housekeeping genes currently used by the research community do in fact change under certain conditions and in different tissues (ClonTechniques 1997b). It is imperative, therefore, that preliminary experiments be carried out on a panel of housekeeping genes to establish their suitability for use in the model system.

Interpretation of quantitative data must also be treated with caution. By comparing the lists of genes identified by differential expression one can perhaps gain insight into why two different species react in different ways to external stimuli. For example, rats and mice appear sensitive to the non-genotoxic effects of a wide range of peroxisome proliferators whilst Syrian hamsters and guinea pigs are largely resistant (Orton *et al.* 1984, Rodricks and Turnbull 1987, Lake *et al.* 1989, 1993, Makowska *et al.* 1992). A simplified approach to resolving the reason(s) why is to compare lists of up- and down-regulated genes in order to identify those which are expressed in only one species and, through background knowledge of the effects of the said gene, might suggest a mechanism of facilitated non-genotoxic carcinogenesis or protection. Of course, the situation is likely to be far more complex. Perhaps if there were one key gene protecting guinea pig from non-genotoxic effects and it was upregulated 50 times by PPs, the same gene might only be up-regulated five times in the rat. However, since both were noted to be upregulated, the importance of the gene may be overlooked. Just to complicate matters, a large change in expression does not necessarily mean a biologically important change. For example, what is the true relevance of gene Y which shows a 50-fold increase after a particular treatment, and gene Z which shows only a 5-fold increase? If one examines the literature one may find that historically, gene Y has often been shown to be up-regulated 40–60-fold by a number of unrelated stimuli—in light of this the 50-fold increase would appear less significant. However, the literature may show that gene Z has never been recorded as having more than doubled in expression—which makes your 5-fold increase all the more exciting. Perhaps even more interesting is if that same 5-fold increase has only been seen in related neoplasms or following treatment with related chemicals.

Problems in using the differential display approach

Differential display technology originally held promise of an easily obtainable 'fingerprint' of those genes which are up- or down-regulated in test animals/cells in a developmental process or following exposure to given stimuli. However, it has

become clear that the fingerprinting process, whilst still valid, is much too complex to be represented by a single technique profile. This is because all differential display techniques have common and/or unique technical problems which preclude the isolation and identification of all those genes which show changes in expression. Furthermore, there are important genetic changes related to disease development which differential expression analysis is simply not designed to address. An example of this is the presence of small deletions, insertions, or point mutations such as those seen in activated oncogenes, tumour suppressor genes and individual polymorphisms. Polymorphic variations, small though they usually are, are often regarded as being of paramount importance in explaining why some patients respond better than others to certain drug treatments (and, in logical extension, why some people are less affected by potentially dangerous xenobiotics/carcinogens than others). The identification of such point mutations and naturally occurring polymorphisms requires the subsequent application of sequencing, SSCP, DGGE or TGGE to the gene of interest. Furthermore, differential display is not designed to address issues such as alternatively spliced gene species or whether an increased abundance of mRNA is a result of increased transcription or increased mRNA stability.

Conclusions

Perhaps the main advantage of open system differential display techniques is that they are not limited by extant theories or researcher bias in revealing genes which are differentially expressed, since they are designed to amplify all genes which demonstrate altered expression. This means that they are useful for the isolation of previously unknown genes which may turn out be useful biomarkers of a particular state or condition. At least one open system (SAGE) is also quantitative, thus eliminating the need to return to the original mRNA and carry out Northern/PCR analysis to confirm the result. However, the rapid progress of genome mapping projects means that over the next 5–10 years or so, the balance of experimental use will switch from open to closed differential display systems, particularly DNA arrays. Arrays are easier and faster to prepare and use, provide quantitative data, are suitable for high throughput analysis and can be tailored to look at specific signalling pathways or families of genes. Identification of all the gene sequences in human and common laboratory animals combined with improved DNA array technology, means that it will soon no longer be necessary to try to isolate differentially expressed genes using the technically more demanding open system approach. Thus, their main advantage (that of identifying unknown genes) will be largely eradicated. It is likely, therefore, that their sphere of application will be reduced to analysis of the less common laboratory species, since it will be some time yet before the genomes of such animals as zebrafish, electric eels, gerbils, crayfish and squid, for example, will be sequenced.

Of course, in the end the question will always remain: What is the functional/biological significance of the identified, differentially expressed genes? One persistent problem is understanding whether differentially expressed genes are a cause or consequence of the altered state. Furthermore, many chemicals, such as non-genotoxic carcinogens, are also mitogens and so genes associated with replication will also be upregulated but may have little or nothing to do with the

carcinogenic effect. Whilst differential display technology cannot hope to answer these questions, it does provide a springboard from which identification, regulatory and functional studies can be launched. Understanding the molecular mechanism of cellular responses is almost impossible without knowing the regulation and function of those genes and their condition (e.g. mutated). In an abstract sense, differential display can be likened to a still photograph, showing details of a fixed moment in time. Consider the Historian who knows the outcome of a battle and the placement and condition of the troops before the battle commenced, but is asked to try and deduce how the battle progressed and why it ended as it did from a few still photographs—an impossible task. In order to understand the battle, the Historian must find out the capabilities and motivation of the soldiers and their commanding officers, what the orders were and whether they were obeyed. He must examine the terrain, the remains of the battle and consider the effects the prevailing weather conditions exerted. Likewise, if mechanistic answers are to be forthcoming, the scientist must use differential display in combination with other techniques, such as knockout technology, the analysis of cell signalling pathways, mutation analysis and time and dose response analyses. Although this review has emphasized the importance of differential gene profiling, it should not be considered in isolation and the full impact of this approach will be strengthened if used in combination with functional genomics and proteomics (2-dimensional protein gels from isoelectric focusing and subsequent SDS electrophoresis and virtual 2D-maps using capillary electrophoresis). Proteomics is attracting much recent attention as many of the changes resulting in differential gene expression do not involve changes in mRNA levels, as decribed extensively herein, but rather protein–protein, protein–DNA and protein phosphorylation events which would require functional genomics or proteomic technologies for investigation.

Despite the limitations of differential display technology, it is clear that many potential applications and benefits can be obtained from characterizing the genetic changes that occur in a cell during normal and disease development and in response to chemical or biological insult. In light of functional data, such profiling will provide a 'fingerprint' of each stage of development or response, and in the long term should help in the elucidation of specific and sensitive biomarkers for different types of chemical/biological exposure and disease states. The potential medical and therapeutic benefits of understanding such molecular changes are almost immeasurable. Amongst other things, such fingerprints could indicate the family or even specific type of chemical an individual has been exposed to plus the length and/or acuteness of that exposure, thus indicating the most prudent treatment. They may also help uncover differences in histologically identical cancers, provide diagnostic tests for the earliest stages of neoplasia and, again, perhaps indicate the most efficacious treatment.

The Human Genome Project will be completed early in the next century and the DNA sequence of all the human genes will be known. The continuing development and evolution of differential gene expression technology will ensure that this knowledge contributes fully to the understanding of human disease processes.

Acknowledgements

We acknowledge Drs Nick Plant (University of Surrey), Sally Darney and Chris Luft (US EPA at RTP) for their critical analysis of the manuscript prior to submission. This manuscript has been reviewed in accordance with the policy of the

US Environmental Protection Agency and approved for publication. Approval does not signify that the contents reflect the views and policies of the Agency, nor does mention of trade names constitute endorsement or recommendation for use.

References

ADAMS, M. D., KELLEY, J. M., GOCAYNE, J. D., DUBNICK, M., POLYMEROPOULOS, M. H., XIAO, H., MERRIL, C. R., WU, A., OLDE, B., MORENO, R. F., KERLAVAGE, A. R., MCCOMBIE, W. R. and VENTOR, J. C., 1991, Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651–1656.

AN, G., LUO, G., VELTRI, R. W. and O'HARA, S. M., 1996, Sensitive non-radioactive differential display method using chemiluminescent detection. *Biotechniques*, **20**, 342–346.

AXEL, R., FEIGELSON, P. and SCHULTZ, G., 1976, Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. *Cell*, **7**, 247–254.

BAND, V. and SAGER, R., 1989, Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proceedings of the Naional Academy of Sciences, USA*, **86**,1249–1253.

BAUER, D., MULLER, H., REICH, J., RIEDEL, H., AHRENKIEL, V., WARTHOE, P. and STRAUSS, M., 1993, Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research*, **21**, 4272–4280.

BERTIOLI, D. J., SCHLICHTER, U. H. A., ADAMS, M. J., BURROWS, P. R., STEINBISS, H.-H. and ANTONIW, J. F., 1995, An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Research*, **23**, 4520–4523.

BRAVO, R., 1990, Genes induced during the G0/G1 transition in mouse fibroblasts. *Seminars in Cancer Biology*, **1**, 37–46.

BURN, T. C., PETROVICK, M. S., HOHAUS, S., ROLLINS, B. J. and TENEN, D. G., 1994, Monocyte chemoattractant protein-1 gene is expressed in activated neutrophils and retinoic acid-induced human myeloid cell lines. *Blood*, **84**, 2776–2783.

CAO, J., CAI, X., ZHENG, L., GENG, L., SHI, Z., PAO, C. C. and ZHENG, S., 1997, Characterisation of colorectal cancer-related cDNA clones obtained by subtractive hybridisation screening. *Journal of Cancer Research and Clinical Oncology*, **123**, 447–451.

CASSIDY, S. B., 1995, Uniparental disomy and genomic imprinting as causes of human genetic disease. *Environmental and Molecular Mutagenesis*, **25** (Suppl 26), 13–20.

CHANG, G. W. and TERZAGHI-HOWE, M., 1998, Multiple changes in gene expression are associated with normal cell-induced modulation of the neoplastic phenotype. *Cancer Research*, **58**, 4445–4452.

CHEN, J., SCHWARTZ, D. A., YOUNG, T. A., NORRIS, J. S. and YAGER, J. D., 1996, Identification of genes whose expression is altered during mitosuppression in livers of ethinyl estradiol-treated female rats. *Carcinogenesis*, **17**, 2783–2786.

CHEN, J. J. W. and PECK, K., 1996, Non-radioactive differential display method to directly visualise and amplify differential bands on nylon membrane. *Nucleic Acid Research*, **24**, 793–794.

CLON TECHNIQUES, 1997a, PCR-Select Differential Screening Kit—the nextstep after Clontech PCR-Select cDNA subtraction. *ClonTechniques*, **XII**, 18–19.

CLON TECHNIQUES, 1997b, Housekeeping RT-PCR amplimers and cDNA probes. *ClonTechniques*, **XII**, 15–16.

DAVIS, M. M., COHEN, D. I., NIELSEN, E. A., STEINMETZ, M., PAUL, W. E. and HOOD, L., 1984, Cell-type-specific cDNA probes and the murine I region: the localization and orientation of Ad alpha. *Proceedings of the National Academy of Sciences (USA)*, **81**, 2194–2198.

DELLAVALLE, R. P., PETERSON, R. and LINDQUIST, S., 1994, Preferential deadenylation of HSP70 mRNA plays a key role in regulating Hsp70 expression in Drosophila melanogaster. *Molecular and Cell Biology*, **14**, 3646–3659.

DERISI, J. L., VASHWANATH, R. L. and BROWN, P., 1997, Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

DIATCHENKO, L., LAU, Y.-F. C., CAMPBELL, A. P., CHENCHIK, A., MOQADAM, F., HUANG, B., LUKYANOV, K., GURSKAYA, N., SVERDLOV, E. D. and SIEBERT, P. D., 1996, Suppression subtractive hybridisation: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences (USA)*, **93**, 6025–6030.

DOGRA, S. C., WHITELAW, M. L. and MAY, B. K., 1998, Transcriptional activation of cytochrome P450 genes by different classes of chemical inducers. *Clinical and Experimental Pharmacology and Physiology*, **25**, 1–9.

DUGUID, J. R. and DINAUER, M. C., 1990, Library subtraction of *in vitro* cDNA libraries to identify differentially expressed genes in scrapie infection. *Nucleic Acids Research*, **18**, 2789–2792.

DUNBAR, P. R., OGG, G. S., CHEN, J., RUST, N., VAN DER BRUGGEN, P. and CERUNDOLO, V., 1998, Direct isolation, phenotyping and cloning of low-frequency antigen-specific cytotoxic T lymphocytes from peripheral blood. *Current Biology*, **26**, 413–416.

FITZPATRICK ,D. R., GERMAIN -LEE, E. and VALLE, D., 1995, Isolation and characterisation of rat and human cDNAs encoding a novel putative peroxisomal enoyl-CoA hydratase. *Genomics*, **27**, 457–466.

FOSS, D. L., BAARSCH, M. J. and MURTAUGH, M. P., 1998, Regulation of hypoxanthine phosphoribosyltransferase, glyceraldehyde-3-phosphate dehydrogenase and beta-actin mRNA expression in porcine immune cells and tissues. *Animal Biotechnology*, **9**, 67–78 .

FRYE, R. A., BENZ, C. C. and LIU, E., 1989, Detection of amplified oncogenes by differential polymerase chain reaction. *Oncogene*, **4**, 1153–1157.

GEISINGER , A., RODRIGUEZ, R., ROMERO , V. and WETTSTEIN R., 1997, A simple method for screening cDNAs arising from the cloning of RNA differential display bands. *Elsevier Trends Journals Technical Tips Online*, http://tto.trends.com, document T01110.

GRESS, T. M., HOHEISEL , J. D., LENNON , G. G., ZEHETNER , G. and LEHRACH , H., 1992, Hybridisation fingerprinting of high density cDNA filter arrays with cDNA pools derived from whole tissues. *Mammalian Genome*, **3**, 609–619.

GRIFFIN , G. and KRISHNA , S., 1998, Cytokines in infectious diseases. *Journal of the Royal College of Physicians, London*, **32**, 195–198.

GROENINK , M. and LEEGWATER , A. C. J., 1996, Isolation of delayed early genes associated with liver regeneration using Clontech PCR-select subtraction technique. *Clontechniques*, **XI**, 23–24.

GUIMARAES , M. J., BAZAN, J. F., ZLOTNIK , A., WILES , M. V., GRIMALDI , J. C., LEE, F. and MCCLANAHAN , T., 1995b, A new approach to the study of haematopoietic development in the yolk sac and embryoid bodies. *Development*, 121, 3335–3346.

GUIMERAES , M. J., LEE, F., ZLOTNIK , A. and MCCLANAHAN , T., 1995a, Differential display by PCR:novel findings and applications. *Nucleic Acids Research*, **23**, 1832–1833.

GURSKAYA, N. G., DIATCHENKO , L., CHENCHIK , P. D., SIEBERT , P. D., KHASPEKOV , G. L., LUKYANOV , K. A., VAGNER, L. L., ERMOLAEVA , O. D., LUKYANOV, S. A. and SVERDLOV , E. D., 1996, Equalising cDNA subtraction based on selective suppression of polymerase chain reaction: Cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-Myrystate 13-Acetate. *Analytical Biochemistry*, **240**, 90–97.

HAMPSON , I. N. and HAMPSON , L., 1997, CCLS and DROP—subtractive cloning made easy. *Life Science News* (A publication of Amersham Life Science), **23**, 22–24.

HAMPSON , I. N., HAMPSON , L. and DEXTER , T. M., 1996, Directional random oligonucleotide primed (DROP) global amplification of cDNA: its application to subtractive cDNA cloning. *Nucleic Acids Research*. **24**, 4832–4835.

HAMPSON , I. N., POPE, L., COWLING , G. J. and DEXTER , T. M., 1992, Chemical cross linking subtraction (CCLS): a new method for the generation of subtractive hybridisation probes. *Nucleic Acids Research*, **20**, 2899.

HARA, E., KATO, T., NAKADA, S., SEKIYA , S. and ODA, K., 1991, Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. *Nucleic Acids Research*, **19**, 7097–7104.

HATADA, I., HAYASHIZAKE , Y., HIROTSUNE , S., KOMATSUBARA , H. and MUKAI, T., 1991, A genomic scanning method for higher organisms using restriction sites as landmarks. *Proceedings of the National Academy of Sciences (USA)*, **88**, 9523–9527.

HECHT, N., 1998, Molecular mechanisms of male sperm cell differentiation. *Bioessays*, **20**, 555–561.

HEDRICK , S., COHEN , D. I., NIELSEN , E. A. and DAVIS , M. E., 1984, Isolation of T cell-specific membrane-associated proteins. *Nature*, **308**, 149–153.

HERTZ, R., SECKBACH , M., ZAKIN , M. M. and BAR-TANA, J., 1996, Transcriptional suppression of the transferrin gene by hypolipidemic peroxisome proliferators. *Journal of Biological Chemistry*, **271**, 218–224.

HEUVAL , J. P. V., CLARK, G. C., KOHN , M. C., TRITSCHER , A. M., GREENLEE , W. F., LUCIER , G. W. and BELL , D. A., 1994, Dioxin-responsive genes: Examination of dose-response relationships using quantitative reverse transciptase-polymerase chain reaction. *Cancer Research*, **54**, 62–68.

HILLIER , L. D., LENNON , G., BECKER , M., BONALDO , M. F., CHIAPELLI , B., CHISSOE , S., DIETRICH , N., DUBUQUE, T., FAVELLO , A., GISH , W., HAWKINS , M., HULTMAN , M., KUCABA, T., LACY, M., LE, M., LE, N., MARDIS, E., MOORE, B., MORRIS , M., PARSONS , J., PRANGE, C., RIFKIN , L., ROHLFING , T., SCHELLENBERG , K., SOARES, M. B., TAN, F., THIERRY -MEG, J., TREVASKIS , E., UNDERWOOD , K., WOHLDMAN , P., WATERSTON , R., WILSON , R and MARRA, M., 1996, Generation and analysis of 280,000 human expressed sequence tags. *Genome Research*, **6**, 807–828.

HUBANK, M. and SCHATZ, D. G., 1994, Identifying differences in mRNA expression by representational difference analysis. *Nucleic Acids Research*, **22**, 5640–5648.

HUNTER , T., 1991, Cooperation between oncogenes. *Cell*, **64**, 249–270.

IVANOVA , N. B. and BELYAVSKY , A. V., 1995, Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting. *Nucleic Acids Research*, **23**, 2954–2958.

JAMES , B. D. and HIGGINS , S. J, 1985, *Nucleic Acid Hybridisation* (Oxford: IRL Press Ltd).

KAS-DEELEN , A. M., HARMSEN , M. C., DE MAAR, E. F. and VAN SON, W. J, 1998, A sensitive method for

quantifying cytomegalic endothelial cells in peripheral blood from cytomegalovirus-infected patients. *Clinical Diagnostic and Laboratory Immunology*, **5**, 622–626.

KILTY, I. and VICKERS, P., 1997, Fractionating DNA fragments generated by differential display PCR. *Strategies Newsletter* (Stratagene), **10**, 50–51.

KLEINJAN, D.-J. and VAN HEYNINGEN, V., 1998, Position effect in human genetic disease. *Human and Molecular Genetics*, **7**, 1611–1618.

KO, M. S., 1990, An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs. *Nucleic Acids Research*, **18**, 5705–5711.

LAKE, B. G., EVANS, J. G., CUNNINGHAME, M. E. and PRICE, R. J., 1993, Comparison of the hepatic effects of Wy-14,643 on peroxisome proliferation and cell replication in the rat and Syrian hamster. *Environmental Health Perspectives*, **101**, 241–248.

LAKE, B. G., EVANS, J. G., GRAY, T. J. B., KOROSI, S. A. and NORTH, C. J., 1989, Comparative studies of nafenopin-induced hepatic peroxisome proliferation in the rat, Syrian hamster, guiea pig and marmoset. *Toxicology and Applied Pharmacology*, **99**, 148–160.

LENNARD, M. S., 1993, Genetically determined adverse drug reactions involving metabolism. *Drug Safety*, **9**, 60–77.

LEVY, S., TODD, S. C. and MAECKER, H. T., 1998, CD81(TAPA-1): a molecule involved in signal transduction and cell adhesion in the immune system. *Annual Review of Immunology*, **16**, 89–109.

LIANG, P. and PARDEE, A. B., 1992, Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967–971.

LIANG, P., AVERBOUKH, L., KEYOMARSI, K., SAGER, R. and PARDEE, A., 1992, Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Research*, **52**, 6966–6968.

LIANG, P., AVERBOUKH, L. and PARDEE, A. B., 1993, Distribution & cloning of eukaryotic mRNAs by means of differential display refinements and optimisation. *Nucleic Acids Research*, **21**, 3269–3275.

LIANG, P., BAUER, D., AVERBOUKH, L., WARTHOE, P., ROHRWILD, M., MULLER, H., STRAUSS, M. and PARDEE, A. B., 1995, Analysis of altered gene expression by differential display. *Methods in Enzymology*, **254**, 304–321.

LINSKENS, M. H., FENG, J., ANDREWS, W. H., ENLOW, B. E., SAATI, S. M., TONKIN, L. A., FUNK, W. D. and VILLEPONTEAU, B., 1995, Cataloging altered gene expression in young and senescent cells using enhanced differential display. *Nucleic Acids Research*, **23**, 3244–3251.

LISITSYN, N., LISIITSYN, N. and WIGLER, M., 1993, Cloning the differences between two complex genomes. *Science*, **259**, 946–951.

LOHMANN, J., SCHICKLE, H. and BOSCH, T. C. G., 1995, REN Display, a rapid and efficient method for non-radioactive differential display and mRNA isolation. *Biotechniques*, **18**, 200–202.

LUNNEY, J. K., 1998, Cytokines orchestrating the immune response. *Reviews in Science and Techology*, **17**, 84–94.

MAKOWSKA, J. M., GIBSON, G. G. and BONNER, F. W., 1992, Species differences in ciprofibrate-induction of hepaic cytochrome P4504A1 and peroxisome proliferation. *Journal of Biochemical Toxicology*, **7**, 183–191.

MALDARELLI, F., XIANG, C., CHAMOUN, G. and ZEICHNER, S. L., 1998, The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. *Virus Research*, **53**, 39–51.

MATHIEU-DAUDE, F., CHENG, R., WELSH, J. and MCCLELLAND, M., 1996, Screening of differentially amplified cDNA products from RNA arbitrarily primed PCR fingerprints using single strand conformation polymorphism (SSCP) gels. *Nucleic Acids Research*, **24**, 1504–1507.

MCKENZIE, D. and DRAKE, D., 1997, Identification of differentially expressed gene products with the castaway system. *Strategies Newsletter* (Stratagene), **10**,19–20.

MCCLELLAND, M., MATHIEU-DAUDE, F. and WELSH, J., 1996, RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends in Genetics*, **11**, 242–246.

MECHLER, B. and RABBITTS, T. H., 1981, Membrane-bound ribosomes of myeloma cells. IV. mRNA complexity of free and membrane-bound polysomes. *Journal of Cell Biology*, **88**, 29–36.

MEYER, U. A. and ZANGER, U. M., 1997, Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annual Review of Pharmacology and Toxicology*, **37**, 269–296.

MOHLER, K. M. and BUTLER, L. D., 1991, Quantitation of cytokine mRNA levels utilizing the reverse transcriptase-polymerase chain reaction following primary antigen-specific sensitization in vivo—I. Verification of linearity, reproducibility and specificity. *Molecular Immunology*, **28**, 437–447.

MURPHY, L. D., HERZOG, C. E., RUDICK, J. B., TITO FOJO, A. and BATES, S. E., 1990, Use of the polymerase chain reaction in the quantitation of the mdr-1 gene expression. *Biochemistry*, **29**, 10351–10356.

NELSON, D. R., KOYMANS, L., KAMATAKI, T., STEGEMAN, J. J., FEYEREISEN, R., WAXMAN, D. J., WATERMAN, M. R., GOTOH, O., COON, M. J., ESTABTROOK, R. W., GUNSALUS, I. C. and NEBERT, D. W., 1996, Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, **6**, 1–42.

NISHIO, Y., AIELLO, L. P. and KING, G. L., 1994, Glucose induced genes in bovine aortic smooth muscle cells identified by mRNA differential display. *FASEB Journal*, **8**, 103–106.

O'NEILL, M. J. and SINCLAIR, A. H., 1997, Isolation of rare transcripts by representational difference analysis. *Nucleic Acids Research*, **25**, 2681–2682.

ORTON, T. C., ADAM, H. K., BENTLEY, M., HOLLOWAY, B. and TUCKER, M. J., 1984, Clobuzarit: species differences in the morphological and biochemical response of the liver following chronic administration. *Toxicology and Applied Pharmacology*, **73**, 138–151.

PELKONEN, O., MAENPAA, J., TAAVITSAINEN, P., RAUTIO, A. and RAUNIO, H., 1998, Inhibition and Induction of human cytochrome P450 (CYP) enzymes. *Xenobiotica*, **28**, 1203–1253.

PHILIPS, S. M., BENDALL, A. J. and RAMSHAW, I. A., 1990, Isolation of genes associated with high metastatic potential in rat mammary adenocarcinomas. *Journal of the National Cancer Institute*, **82**, 199–203.

PRASHAR, Y. and WEISSMAN, S. M., 1996, Analysis of differential gene expression by display of 3'end restriction fragments of cDNAs. *Proceedings of the National Academy of Sciences (USA)*, **93**, 659–663.

RAGNO, S., ESTRADA, I., BUTLER, R. and COLSTON, M. J., 1997, Regulation of macrophage gene expression following invasion by *Mycobacterium tuberculosis*. *Immunology Letters*, **57**, 143–146.

RAMANA, K. V. and KOHLI, K. K., 1998, Gene regulation of cytochrome P450—an overview. *Indian Journal of Experimental Biology*, **36**, 437–446.

RICHARD, L., VELASCO, P. and DETMAR, M., 1998, A simple immunomagnetic protocol for the selective isolation and long-term culture of human dermal microvascular endothelial cells. *Experimental Cell Research*, **240**, 1–6.

ROCKETT, J. C., ESDAILE, D. J. and GIBSON, G. G., 1997, Molecular profiling of non-genotoxic hepatocarcinogenesis using differential display reverse transcription-polymerase chain reaction (ddRT-PCR). *European Journal of Drug. Metabolism and Pharmacokinetics*, **22**, 329–333.

RODRICKS, J. V. and TURNBULL, D., 1987, Inter-species differences in peroxisomes and peroxisome proliferation. *Toxicology and Industrial Health*, **3**, 197–212.

ROGLER, G., HAUSMANN, M., VOGL, D., ASCHENBRENNER, E., ANDUS, T., FALK, W., ANDREESEN, R., SCHOLMERICH, J. and GROSS, V., 1998, Isolation and phenotypic characterization of colonic macrophages. *Clinical and Experimental Immunology*, **112**, 205–215.

ROHN, W. M., LEE, Y. J. and BENVENISTE, E. N., 1996, Regulation of class II MHC expression. *Critical Reviews in Immunology*, **16**, 311–330.

RUDIN, C. M. and THOMPSON, C. B., 1998, B-cell development and maturation. *Seminars in Oncology*, **25**, 435–446.

SAKAGUCHI, N., BERGER, C. N. and MELCHERS, F., 1986, Isolation of a cDNA copy of an RNA species expressed in murine pre-B cells. *EMBO Journal*, **5**, 2139–2147.

SAMBROOK, J., FRITSCH, E. F. and MANIATIS, T., 1989, Gel electrophoresis of DNA. In N. Ford, M. Nolan and M. Fergusen (eds), *Molecular Cloning—A laboratory manual*, 2nd edition (New York: Cold Spring Harbour Laboratory Press), Volume 1, pp. 6–37.

SARGENT, T. D. and DAWID, I. B., 1983, Differential gene expression in the gastrula of Xenopus laevis. *Science*, **222**, 135–139.

SCHENA, M., SHALON, D., HELLER, R., CHAI, A., BROWN., P. O. and DAVIS, R. W., 1996, Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences (USA)*, **93**, 10614–10619.

SCHNEIDER, C., KING, R. M. and PHILIPSON, L., 1988, Genes specifically expressed at growth arrest of mammalian cells. *Cell*, **54**, 787–793.

SCHNEIDER-MAUNOURY, S., GILARDI-HEBENSTREIT, P. and CHARNAY, P., 1998, How to build a vertebrate hindbrain. Lessons from genetics. *C R Academy of Science III*, **321**, 819–834.

SEMENZA, G. L., 1994, Transcriptional regulation of gene expression: mechanisms and pathophysiology. *Human Mutations*, **3**, 180–199.

SEWALL, C. H., BELL, D. A., CLARK, G. C., TRITSCHER, A. M., TULLY, D. B., VANDEN HEUVEL, J. and LUCIER, G. W., 1995, Induced gene transcription: implications for biomarkers. *Clinical Chemistry*, **41**, 1829–1834.

SINGH, N., AGRAWAL, S. and RASTOGI, A. K., 1997, Infectious diseases and immunity: special reference to major histocompatibility complex. *Emerging Infectious Diseases*, **3**, 41–49.

SMITH, N. R., LI, A., ALDERSLEY, M., HIGH, A. S., MARKHAM, A. F. and ROBINSON, P. A., 1997, Rapid determination of the complexity of cDNA bands extracted from DDRT-PCR polyacrylamide gels. *Nucleic Acids* Research, **25**, 3552–3554.

SOMPAYRAC, L., JANE, S., BURN., T. C., TENEN, D. G. and DANNA, K. J., 1995, Overcoming limitations of the mRNA differential display technique. *Nucleic Acids Research*, **23**, 4738–4739.

ST JOHN, T. P. and DAVIS, R. W., 1979, Isolation of galactose-inducible DNA sequences from Saccharomyces cerevisiae by differential plaque filter hybridisation. *Cell*, **16**, 443–452.

SUN, Y., HEGAMYER, G. and COLBURN, N. H., 1994, Molecular cloning of five messenger RNAs differentially expressed in preneoplastic or neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinases-3. *Cancer Research*, **54**, 1139–1144.

SUNG, Y. J. and DENMAN, R. B., 1997, Use of two reverse transcriptases eliminates false-positive results in differential display. *Biotechniques*, **23**, 462–464.

SUTTON, G., WHITE, O., ADAMS, M. and KERLAVAGE, A., 1995, TIGR Assembler; A new tool for assembling large shotgun sequencing projects. *Genome Science and Technology*, **1**, 9–19.

SUZUKI, Y., SEKIYA, T. and HAYASHI, K., 1991, Allele-specific polymerase chain reaction: a method for amplification and sequence determination of a single component among a mixture of sequence variants. *Analytical Biochemistry*, **192**, 82–84.

SYED, V., GU, W. and HECHT, N. B., 1997, Sertoli cells in culture and mRNA differential display provide a sensitive early warning assay system to detect changes induced by xenobiotics. *Journal of Andrology*, **18**, 264–273.

UITTERLINDEN, A. G., SLAGBOOM, P., KNOOK, D. L. and VIJGL, J., 1989, Two-dimensional DNA fingerprinting of human individuals. *Proceedings of the National Academy of Sciences (USA)*, **86**, 2742–2746.

ULLMAN, K. S., NORTHROP, J. P., VERWEIJ, C. L. and CRABTREE, G. R., 1990, Transmission of signals from the T lymphocyte antigen receptor to the genes responsible for cell proliferation and immune function: the missing link. *Annual Review of Immunology*, **8**, 421–452.

VASMATZIS, G., ESSAND, M., BRINKMANN, U., LEE, B. and PASTON, I., 1998, Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proceedings of the National Academy of Sciences (USA)*, **95**, 300–304.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. and KINZLER, K. W., 1995, Serial analysis of gene expression. *Science*, **270**, 484–487.

VOELTZ, G. K. and STEITZ, J. A., 1998, AuuuA sequences direct mRNA deadenylation uncoupled from decay during Xenopus early development. *Molecular and Cell Biology*, **18**, 7537–7545.

VOGELSTEIN, B. and KINZLER, K. W., 1993, The multistep nature of cancer. *Trends in Genetics*, **9**, 138–141.

WALTER, J., BELFIELD, M., HAMPSON, I. and READ, C., 1997, A novel approach for generating subtractive probes for differential screening by CCLS. *Life Science News*, **21**, 13–14.

WAN, J. S., SHARP, S. J., POIRIER, G. M.-C., WAGAMAN, P. C., CHAMBERS, J., PYATI, J., HOM, Y.-L., GALINDO, J. E., HUVAR, A., PETERSON, P. A., JACKSON, M. R. and ERLANDER, M. G., 1996, Cloning differentially expressed mRNAs. *Nature Biotechnology*, **14**, 1685–1691.

WALTER, J., BELFIELD, M., HAMPSON, I. and READ, C., 1997, A novel approach for generating subtractive probes for differential screening by CCLS, *Life Science News*, **21**, 13–14.

WANG, Z. and BROWN, D. D. 1991, A gene expression screen. *Proceedings of the National Academy of Sciences (USA)*, **88**, 11505–11509.

WAWER, C., RUGGEBERG, H., MEYER, G. and MUYZER, G., 1995, A simple and rapid electrophoresis method to detect sequence variation in PCR-amplified DNA fragments. *Nucleic Acids Research*, **23**, 4928–4929.

WELSH, J., CHADA, K., DALAL, S. S., CHENG, R., RALPH, D. and MCCLELLAND, M., 1992, Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Research*, **20**, 4965–4970.

WONG, H., ANDERSON, W. D., CHENG, T. and RIABOWOL, K. T., 1994, Monitoring mRNA expression by polymerase chain reaction: the 'primer-dropping' method. *Analytical Biochemistry*, **223**, 251–258.

WONG, K. K. and MCCLELLAND, M., 1994, Stress-inducible gene of Salmonella typhimurium identified by arbitrarily primed PCR of RNA. *Proceedings of the National Academy of Sciences (USA)*, **91**, 639–643.

WYNFORD-THOMAS, D., 1991, Oncogenes and anti-oncogenes; the molecular basis of tumour behaviour. *Journal of Pathology*, **165**, 187–201.

XHU, D., CHAN, W. L., LEUNG, B. P., HUANG, F. P., WHEELER, R., PIEDRAFITA, D., ROBINSON, J. H. and LIEW, F. Y., 1998, Selective expression of a stable cell surface molecule on type 2 but not type 1 helper T cells. *Journal of Experimental Medicine*, **187**, 787–794.

YANG, M. and SYTOWSKI, A. J., 1996, Cloning differentially expressed genes by linker capture subtraction. *Analytical Biochemistry*, **237**, 109–114.

ZHAO, N., HASHIDA, H., TAKAHASHI, N., MISUMI, Y. and SAKAKI, Y., 1995, High-density cDNA filter analysis: a novel approach for large scale quantitative analysis of gene expression. *Gene*, **156**, 207–213.

ZHAO, X. J., NEWSOME, J. T. and CIHLAR, R. L., 1998, Up-regulation of two *candida albicans* genes in the rat model of oral candidiasis detected by differential display. *Microbial Pathogenesis*, **25**, 121–129.

ZIMMERMANN, C. R., ORR, W. C., LECLERC, R. F., BARNARD, C. and TIMBERLAKE, W. E., 1980, Molecular cloning and selection of genes regulated in *Aspergillus* development. *Cell*, **21**, 709–715.

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 8945–8947, August 1997
Applied Biological Sciences

Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR

DEVAL A. LASHKARI*†, JOHN H. MCCUSKER‡, AND RONALD W. DAVIS*§

*Departments of Genetics and Biochemistry, Beckman Center, Stanford University, Stanford, CA 94305; and ‡Department of Microbiology, 3020 Duke University Medical Center, Durham, NC 27710

Contributed by Ronald W. Davis, May 20, 1997

ABSTRACT **The recent ability to sequence whole genomes allows ready access to all genetic material. The approaches outlined here allow automated analysis of sequence for the synthesis of optimal primers in an automated multiplex oligonucleotide synthesizer (AMOS). The efficiency is such that all ORFs for an organism can be amplified by PCR. The resulting amplicons can be used directly in the construction of DNA arrays or can be cloned for a large variety of functional analyses. These tools allow a replacement of single-gene analysis with a highly efficient whole-genome analysis.**

The genome sequencing projects have generated and will continue to generate enormous amounts of sequence data. The genomes of *Saccharomyces cerevisiae*, *Escherichia coli*, *Haemophilus influenzae* (1), *Mycoplasma genitalium* (2), and *Methanococcus jannaschii* (3) have been completely sequenced. Other model organisms have had substantial portions of their genomes sequenced as well, including the nematode *Caenorhabditis elegans* (4) and the small flowering plant *Arabidopsis thaliana* (5). This massive and increasing amount of sequence information allows the development of novel experimental approaches to identify gene function.

One standard use of genome sequence data is to attempt to identify the functions of predicted open reading frames (ORFs) within the genome by comparison to genes of known function. Such a comparative analysis of all ORFs to existing sequence data is fast, simple, and requires no experimentation and is therefore a reasonable first step. While finding sequence homologies/motifs is not a substitute for experimentation, noting the presence of sequence homology and/or sequence motifs can be a useful first step in finding interesting genes, in designing experiments and, in some cases, predicting function. However, this type of analysis is frequently uninformative. For example, over one-half of new ORFs in *S. cerevisiae* have no known function (6). If this is the case in a well studied organism such as yeast, the problem will be even worse in organisms that are less well studied or less manipulable. A large, experimentally determined gene function database would make homology/motif searches much more useful.

Experimental analysis must be performed to thoroughly understand the biological function of a gene product. Scaling up from classical "cottage industry" one-gene-oriented approaches to whole-genome analysis would be very expensive and laborious. It is clear that novel strategies are necessary to efficiently pursue the next phase of the genome projects—whole-genome experimental analysis to explore gene expression, gene product function, and other genome functions. Model organisms, such as *S. cerevisiae*, will be extremely important in the development of novel whole-genome analysis techniques and, subsequently, in improving our understanding of other more complex and less manipulable organisms.

The genome sequence can be systematically used as a tool to understand ORFs, gene product function, and other genome regions. Toward this end, a directed strategy has been developed for exploiting sequence information as a means of providing information about biological function (Fig. 1). Efforts have been directed toward the amplification of each predicted ORF or any other region of the genome ranging from a few base pairs to several kilobase pairs. There are many uses for these amplicons—they can be cloned into standard vectors or specialized expression vectors, or can be cloned into other specialized vectors such as those used for two-hybrid analysis. The amplicons can also be used directly by, for example, arraying onto glass for expression analysis, for DNA binding assays, or for any direct DNA assay (7). As a pilot study, synthetic primers were made on the 96-well automated multiplex oligonucleotide synthesizer (AMOS) instrument (8) (Fig. 2). These oligonucleotides were used to amplify each ORF on yeast chromosome V. The current version of this instrument can synthesize three plates of 96 oligonucleotides each (25 bases) in an 8-hr day. The amplification of the entire set of PCR products was then analyzed by gel electrophoresis (Fig. 3). Successful amplification of the proper length product on the first attempt was 95%. This project demonstrates that one can go directly from sequence information to biological analysis in a truly automated, totally directed manner.

These amplicons can be incorporated directly in arrays or the amplicons can be cloned. If the amplicons are to be cloned, novel sequences can be incorporated at the 5′ end of the oligonucleotide to facilitate cloning. One potential problem with cloning PCR products is that the cloned amplicons may contain sequence alterations that diminish their utility. One option would be to resequence each individual amplicon. However, this is expensive, inefficient, and time consuming. A faster, more cost-effective, and more accurate approach is to apply comparative sequencing by denaturing HPLC (9). This method is capable of detecting a single base change in a 2-kb heteroduplex. Longer amplicons can be analyzed by use of appropriate restriction fragments. If any change is detected in a clone, an alternate clone of the same region can be analyzed. Modifying the system to allow high throughput analysis by denaturing HPLC is also relatively simple and straightforward.

If amplicons are used directly on arrays without cloning, it is important to note that, even if single PCR product bands are observed on gels, the PCR products will be contaminated with various amounts of other sequences. This contamination has the potential to affect the results in, for example, expression

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Synteni, Inc., 6519 Dumbarton Circle, Fremont, CA 94555.
§To whom reprint requests should be addressed at: Department of Biochemistry, Beckman Center, B400, Stanford University, Stanford, CA 94305-5307. e-mail: gilbert@cmgm.stanford.edu.

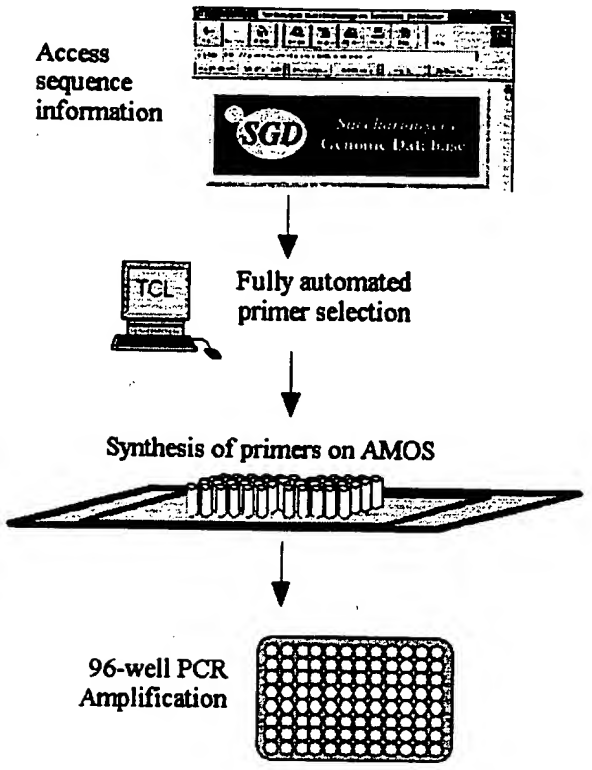


FIG. 1. Overview of systematic method for isolating individual genes. Sequence information is obtained automatically from sequence databases. The data are input into primer selection software specifically designed to target ORFs as designated by database annotations. The output file containing the primer information is directly read by a high-throughput oligonucleotide synthesizer, which makes the oligonucleotides in 96-well plates (AMOS, automated multiplex oligonucleotide synthesizer). The forward and reverse primers are synthesized in the same location on separate plates to facilitate the downstream handling of primers. The amplicons are generated by PCR in 96-well plates as well.

analysis. On the other hand, direct use of the amplicons is much less labor intensive and greatly decreases the occurrence of mistakes in clone identification, a ubiquitous problem associated with large clone set archiving and retrieving.

Any large-scale effort to capture each ORF within a genome must rely on automation if cost is to be minimized while efficiency is maximized. Toward that end, primers targeting ORFs were designed automatically using simple new scripts and existing primer selection software. These script-selected primer sequences were directly read by the high-throughput synthesizer and the forward and reverse primers were synthesized in separate plates in corresponding wells to facilitate automated pipetting and PCR amplifications. Each of the resulting PCR products, generated with minimum labor, contains a known, unique ORF.

Large-scale genome analysis projects are dependent on newly emerging technologies to make the studies practical and economically feasible. For example, the cost of the primers, a significant issue in the past, has been reduced dramatically to make feasible this and other projects that require tens of thousands of oligonucleotides. Other methods of high-throughput analysis are also vital to the success of functional analysis projects, such as microarraying and oligonucleotide chip methods (10–14).

Changes in attitude are also required. One of the major costs of commercial oligonucleotides is extensive quality control such that virtually 100% of the supplied oligonucleotides are successfully synthesized and work for their intended purpose.

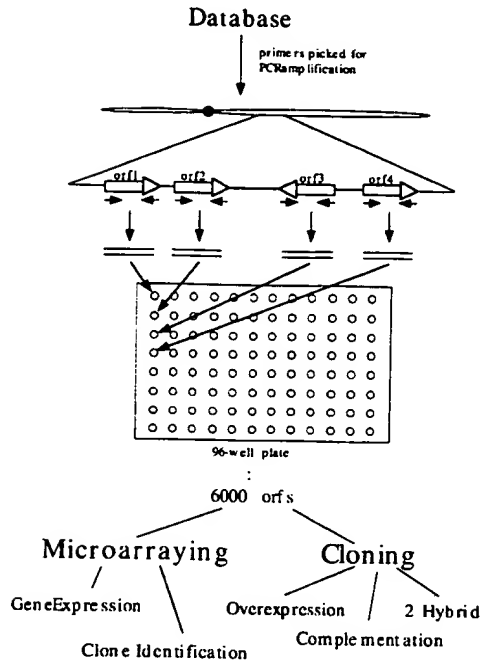


FIG. 2. Overall approach for using database of a genome to direct biological analysis. The synthesis of the 6,000 ORFs (orfs) for each gene of *S. cerevisiae* can be used in many applications utilizing both cloning and microarraying technology.

Considerable cost reduction can be obtained by simply decreasing the expected successful synthesis rate to 95–97%. One can then achieve faster and cheaper whole genome coverage by simply adding a single quality control at the end of the experiment and batching the failures for resynthesis.

The directed nature of the amplicon approach is of clear advantage. The sequence of each ORF is analyzed automatically, and unique specific primers are made to target each ORF. Thus, there is relatively little time or labor involved—for example, no random cloning and subsequent screening is required because each product is known. In the test system, primers for 240 ORFs from chromosome V were systematically synthesized, beginning from the left arm and continuing through to the right arm. At no point was there any manual analysis of sequence information to generate the collection. In many ways, now that the sequence is known, there is no need for the researcher to examine it.

These amplicons can be arrayed and expression analysis can be done on all arrayed ORFs with a single hybridization (10). Those ORFs that display significant differential expression patterns under a given selection are easily identified without the laborious task of searching for and then sequencing a clone. Once scaled up, the procedure provides even greater returns on effort, because a single hybridization will ultimately provide a "snapshot" of the expression of all genes in the yeast genome. Thus, the limiting factor in whole genome analysis will not be the analysis process itself, but will instead be the ability of researchers to design and carry out experimental selections.

Current expression and genetic analysis technologies are geared toward the analysis of single genes and are ill suited to analyze numerous genes under many conditions. Additional difficulties with current technologies include: the effort and expense required to analyze expression and make mutants, the potential duplication of effort if done by different laboratories, and the possibility of conflicting results obtained from different laboratories. In contrast, whole genome analysis not only is more efficient, it also provides data of much higher quality; all genes are assayed and compared in parallel under exactly

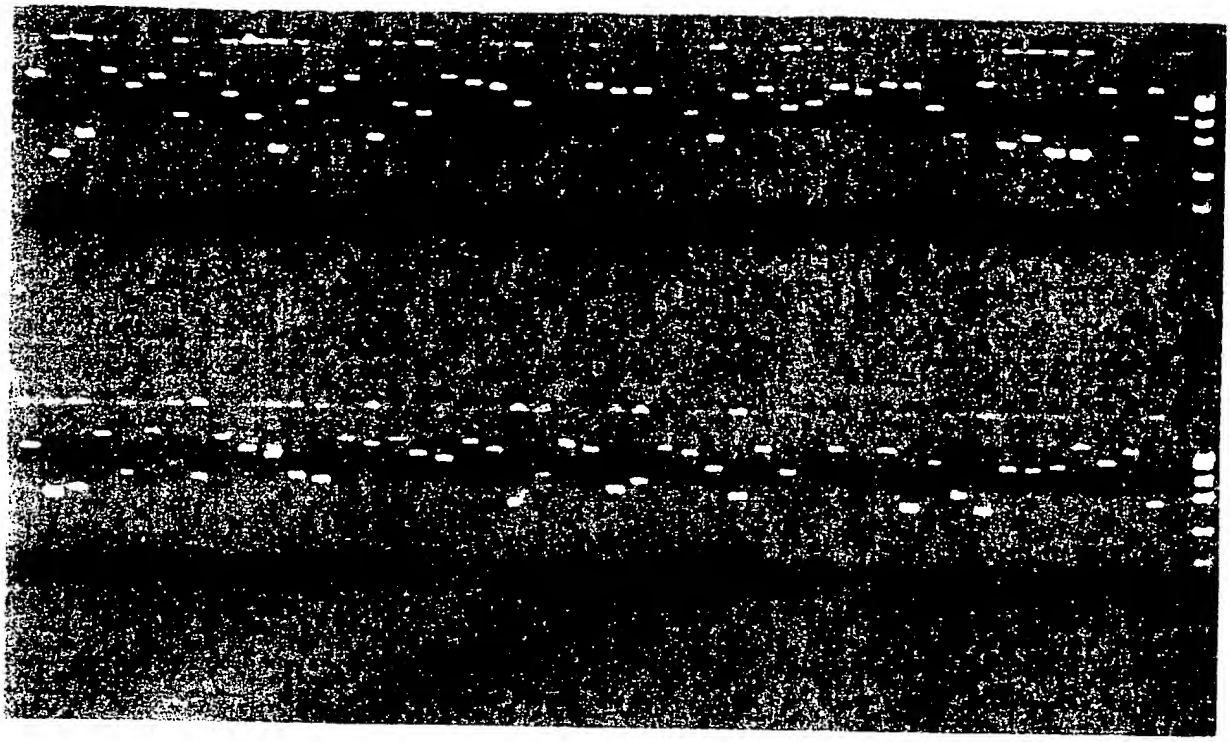


FIG. 3. Gel image of amplifications. Using the method described in Fig. 1, amplicons were generated for ORFs of *S. cerevisiae* chromosome V. One plate of 96 amplification reactions is shown.

the same conditions. In addition, amplicons have many applications beyond gene expression. For example, one recent approach is to incorporate a unique DNA sequence tag, synthesized as part of each gene specific primer, during amplification. The tags or molecular bar codes, when reintroduced into the organism as a gene deletion or as a gene clone, can be used much more efficiently than individual mutations or clones because pools of tagged mutants or transformants can be analyzed in parallel. This parallel analysis is possible because the tags are readily and quantitatively amplified even in complex mixtures of tags (13).

These ORF genome arrays and oligonucleotide tagged libraries can be used for many applications. Any conventional selection applied to a library that gives discrete or multiple products can use these technologies for a simple direct readout. These include screens and selections for mutant complementation, overexpression suppression (15, 16), second-site suppressors, synthetic lethality, drug target overexpression (17), two-hybrid screens (18), genome mismatch scanning (19), or recombination mapping.

The genome projects have provided researchers with a vast amount of information. These data must be used efficiently and systematically to gain a truly comprehensive understanding of gene function and, more broadly, of the entire genome which can then be applied to other organisms. Such global approaches are essential if we are to gain an understanding of the living cell. This understanding should come from the viewpoint of the integration of complex regulatory networks, the individual roles and interactions of thousands of functional gene products, and the effect of environmental changes on both gene regulatory networks and the roles of all gene products. The time has come to switch from the analysis of a single gene to the analysis of the whole genome.

Support was provided by National Institutes of Health Grants R37H60198 and P01H600205.

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., *et al.* (1995) *Science* **269,** 496–512.
2. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., *et al.* (1995) *Science* **270,** 397–403.
3. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273,** 1058–1073.
4. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. (1992) *Nature (London)* **356,** 37–41.
5. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., *et al.* (1994) *Plant Physiol.* **106,** 1241–1255.
6. Oliver, S. (1996) *Nature (London)* **379,** 597–600.
7. Lashkari, D. A. (1996) Ph.D. dissertation (Stanford Univ., Stanford, CA).
8. Lashkari, D. A., Hunicke-Smith, S. P., Norgren, R. M., Davis, R. W. & Brennan, T. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7912–7915.
9. Oefner, P. J. & Underhill, P. A. (1995) *Am. J. Hum. Genet.* **57,** A266.
10. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270,** 467–470.
11. Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* **251,** 767–773.
12. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. (1996) *Science* **274,** 610–614.
13. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. (1996) *Nat. Genet.* **14,** 450–456.
14. Smith, V., Chou, K., Lashkari, D., Botstein, D. & Brown, P. O. (1996) *Science* **274,** 2069–2074.
15. Magdolen, V., Drubin, D. G., Mages, G. & Bandlow, W. (1993) *FEBS Lett.* **316,** 41–47.
16. Ramer, S. W., Elledge, S. J. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11589–11593.
17. Rine, J., Hansen, W., Hardeman, E. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6750–6754.
18. Fields, S. & Song, O. (1989) *Nature (London)* **340,** 245–246.
19. Nelson, S. F., McCusker, J. H., Sander, M. A., Kee, Y., Modrich, P. & Brown, P. O. (1994) *Nat. Genet.* **4,** 11–18.

MOLECULAR CARCINOGENESIS 24:153–159 (1999)

IN PERSPECTIVE
Claudio J. Conti, Editor

Microarrays and Toxicology: The Advent of Toxicogenomics

Emile F. Nuwaysir,[1] Michael Bittner,[2] Jeffrey Trent,[2] J. Carl Barrett,[1] and Cynthia A. Afshari[1]

[1]*Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina*
[2]*Laboratory of Cancer Genetics, National Human Genome Research Institute, Bethesda, Maryland*

The availability of genome-scale DNA sequence information and reagents has radically altered life-science research. This revolution has led to the development of a new scientific subdiscipline derived from a combination of the fields of toxicology and genomics. This subdiscipline, termed toxicogenomics, is concerned with the identification of potential human and environmental toxicants, and their putative mechanisms of action, through the use of genomics resources. One such resource is DNA microarrays or "chips," which allow the monitoring of the expression levels of thousands of genes simultaneously. Here we propose a general method by which gene expression, as measured by cDNA microarrays, can be used as a highly sensitive and informative marker for toxicity. Our purpose is to acquaint the reader with the development and current state of microarray technology and to present our view of the usefulness of microarrays to the field of toxicology. *Mol. Carcinog. 24:153–159, 1999.* © 1999 Wiley-Liss, Inc.

Key words: toxicology; gene expression; animal bioassay

INTRODUCTION

Technological advancements combined with intensive DNA sequencing efforts have generated an enormous database of sequence information over the past decade. To date, more than 3 million sequences, totaling over 2.2 billion bases [1], are contained within the GenBank database, which includes the complete sequences of 19 different organisms [2]. The first complete sequence of a free-living organism, *Haemophilus influenzae,* was reported in 1995 [3] and was followed shortly thereafter by the first complete sequence of a eukaryote, *Saccharomyces cervisiae* [4]. The development of dramatically improved sequencing methodologies promises that complete elucidation of the *Homo sapiens* DNA sequence is not far behind [5].

To exploit more fully the wealth of new sequence information, it was necessary to develop novel methods for the high-throughput or parallel monitoring of gene expression. Established methods such as northern blotting, RNAse protection assays, S1 nuclease analysis, plaque hybridization, and slot blots do not provide sufficient throughput to effectively utilize the new genomics resources. Newer methods such as differential display [6], high-density filter hybridization [7,8], serial analysis of gene expression [9], and cDNA- and oligonucleotide-based microarray "chip" hybridization [10–12] are possible solutions to this bottleneck. It is our belief that the microarray approach, which allows the monitoring of expression levels of thousands of genes simultaneously, is a tool of unprecedented power for use in toxicology studies.

Almost without exception, gene expression is altered during toxicity, as either a direct or indirect result of toxicant exposure. The challenge facing toxicologists is to define, under a given set of experimental conditions, the characteristic and specific pattern of gene expression elicited by a given toxicant. Microarray technology offers an ideal platform for this type of analysis and could be the foundation for a fundamentally new approach to toxicology testing.

MICROARRAY DEVELOPMENT AND APPLICATIONS

cDNA Microarrays

In the past several years, numerous systems were developed for the construction of large-scale DNA arrays. All of these platforms are based on cDNAs or oligonucleotides immobilized to a solid support. In the cDNA approach, cDNA (or genomic) clones of interest are arrayed in a multi-well format and amplified by polymerase chain reaction. The products of this amplification, which are usually 500- to 2000-bp clones from the 3′ regions of the genes of interest, are then spotted onto solid support by using high-speed robotics. By using this method, microarrays of up to 10 000 clones can be generated by spotting onto a glass substrate

*Correspondence to: Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, 111 Alexander Drive, Research Triangle Park, NC 27709.

Received 8 December 1998; Accepted 5 January 1999

Abbreviations: PAH, polycyclic aromatic hydrocarbon; NIEHS, National Institute of Environmental Health Sciences.

[13,14]. Sample detection for microarrays on glass involves the use of probes labeled with fluorescent or radioactive nucleotides.

Fluorescent cDNA probes are generated from control and test RNA samples in single-round reverse-transcription reactions in the presence of fluorescently tagged dUTP (e.g., Cy3-dUTP and Cy5-dUTP), which produces control and test products labeled with different fluors. The cDNAs generated from these two populations, collectively termed the "probe," are then mixed and hybridized to the array under a glass coverslip [10,11,15]. The fluorescent signal is detected by using a custom-designed scanning confocal microscope equipped with a motorized stage and lasers for fluor excitation [10,11,15]. The data are analyzed with custom digital image analysis software that determines for each DNA feature the ratio of fluor 1 to fluor 2, corrected for local background [16,17]. The strength of this approach lies in the ability to label RNAs from control and treated samples with different fluorescent nucleotides, allowing for the simultaneous hybridization and detection of both populations on one microarray. This method eliminates the need to control for hybridization between arrays. The research groups of Drs. Patrick Brown and Ron Davis at Stanford University spearheaded the effort to develop this approach, which has been successfully applied to studies of *Arabidopsis thaliana* RNA [10], yeast genomic DNA [15], tumorigenic versus non-tumorigenic human tumor cell lines [11], human T-cells [18], yeast RNA [19], and human inflammatory disease–related genes [20]. The most dramatic result of this effort was the first published account of gene expression of an entire genome, that of the yeast *Saccharomyces cervisiae* [21].

In an alternative approach, large numbers of cDNA clones can be spotted onto a membrane support, albeit at a lower density [7,22]. This method is useful for expression profiling and large-scale screening and mapping of genomic or cDNA clones [7,22–24]. In expression profiling on filter membranes, two different membranes are used simultaneously for control and test RNA hybridizations, or a single membrane is stripped and reprobed. The signal is detected by using radioactive nucleotides and visualized by phosphorimager analysis or autoradiography. Numerous companies now sell such cDNA membranes and software to analyze the image data [25–27].

Oligonucleotide Microarrays

Oligonucleotide microarrays are constructed either by spotting prefabricated oligos on a glass support [13] or by the more elegant method of direct in situ oligo synthesis on the glass surface by photolithography [28–30]. The strength of this approach lies in its ability to discriminate DNA molecules based on single base-pair difference. This allows the application of this method to the fields of medical diagnostics, pharmacogenetics, and sequencing by hybridization as well as gene-expression analysis.

Fabrication of oligonucleotide chips by photolithography is theoretically simple but technically complex [29,30]. The light from a high-intensity mercury lamp is directed through a photolithographic mask onto the silica surface, resulting in deprotection of the terminal nucleotides in the illuminated regions. The entire chip is then reacted with the desired free nucleotide, resulting in selected chain elongation. This process requires only 4n cycles (where n = oligonucleotide length in bases) to synthesize a vast number of unique oligos, the total number of which is limited only by the complexity of the photolithographic mask and the chip size [29,31,32].

Sample preparation involves the generation of double-stranded cDNA from cellular poly(A)+ RNA followed by antisense RNA synthesis in an in vitro transcription reaction with biotinylated or fluor-tagged nucleotides. The RNA probe is then fragmented to facilitate hybridization. If the indirect visualization method is used, the chips are incubated with fluor-linked streptavidin (e.g., phycoerythrin) after hybridization [12,33]. The signal is detected with a custom confocal scanner [34]. This method has been applied successfully to the mapping of genomic library clones [35], to de novo sequencing by hybridization [28,36], and to evolutionary sequence comparison of the *BRCA1* gene [37]. In addition, mutations in the cystic fibrosis [38] and BRCA1 [39] gene products and polymorphisms in the human immunodeficiency virus-1 clade B protease gene [40] have been detected by this method. Oligonucleotide chips are also useful for expression monitoring [33] as has been demonstrated by the simultaneous evaluation of gene-expression patterns in nearly all open reading frames of the yeast strain *S. cerevisiae* [12]. More recently, oligonucleotide chips have been used to help identify single nucleotide polymorphisms in the human [41] and yeast [42] genomes.

THE USE OF MICROARRAYS IN TOXICOLOGY

Screening for Mechanism of Action

The field of toxicology uses numerous in vivo model systems, including the rat, mouse, and rabbit, to assess potential toxicity and these bioassays are the mainstay of toxicology testing. However, in the past several decades, a plethora of in vitro techniques have been developed to measure toxicity, many of which measure toxicant-induced DNA damage. Examples of these assays include the Ames test, the Syrian hamster embryo cell transformation assay, micronucleus assays, measurements of sister chromatid exchange and unscheduled DNA synthesis, and many others. Fundamental to all of these methods is the fact that toxicity is often preceded by, and results in, alterations in gene expression. In many cases, these changes in gene expression are a

far more sensitive, characteristic, and measurable endpoint than the toxicity itself. We therefore propose that a method based on measurements of the genome-wide gene expression pattern of an organism after toxicant exposure is fundamentally informative and complements the established methods described above.

We are developing a method by which toxicants can be identified and their putative mechanisms of action determined by using toxicant-induced gene expression profiles. In this method, in one or more defined model systems, dose and time-course parameters are established for a series of toxicants within a given prototypic class (e.g., polycyclic aromatic hydrocarbons (PAHs)). Cells are then treated with these agents at a fixed toxicity level (as measured by cell survival), RNA is harvested, and toxicant-induced gene expression changes are assessed by hybridization to a cDNA microarray chip (Figure 1). We have developed a custom DNA chip, called ToxChip v1.0, specifically for this purpose and will discuss it in more detail below. The changes in gene expression induced by the test agents in the model systems are analyzed, and the common set of changes unique to that class of toxicants, termed a toxicant signature, is determined.

This signature is derived by ranking across all experiments the gene-expression data based on relative fold induction or suppression of genes in treated samples versus untreated controls and selecting the most consistently different signals across the sample set. A different signature may be established for each prototypic toxicant class. Once the signatures are determined, gene-expression profiles induced by unknown agents in these same model systems can then be compared with the established signatures. A match assigns a putative mechanism of action to the test compound. Figure 2 illustrates this signature method for different types of oxidant stressors, PAHs, and peroxisome proliferators. In this example, the unknown compound in question had a gene-expression profile similar to that of the oxidant stressors in the database. We anticipate that this general method will also reveal cross talk between different pathways induced by a single agent (e.g., reveal that a compound has both PAH-like and oxidant-like properties). In the future, it may be necessary to distinguish very subtle differences between compounds within a very large sample set (e.g., thousands of highly similar structural isomers in a combinatorial chemistry library or peptide library). To generate these highly refined signatures, standard statistical clustering techniques or principal-component analysis can be used.

For the studies outlined in Figure 2, we developed the custom cDNA microarray chip ToxChip v1.0.

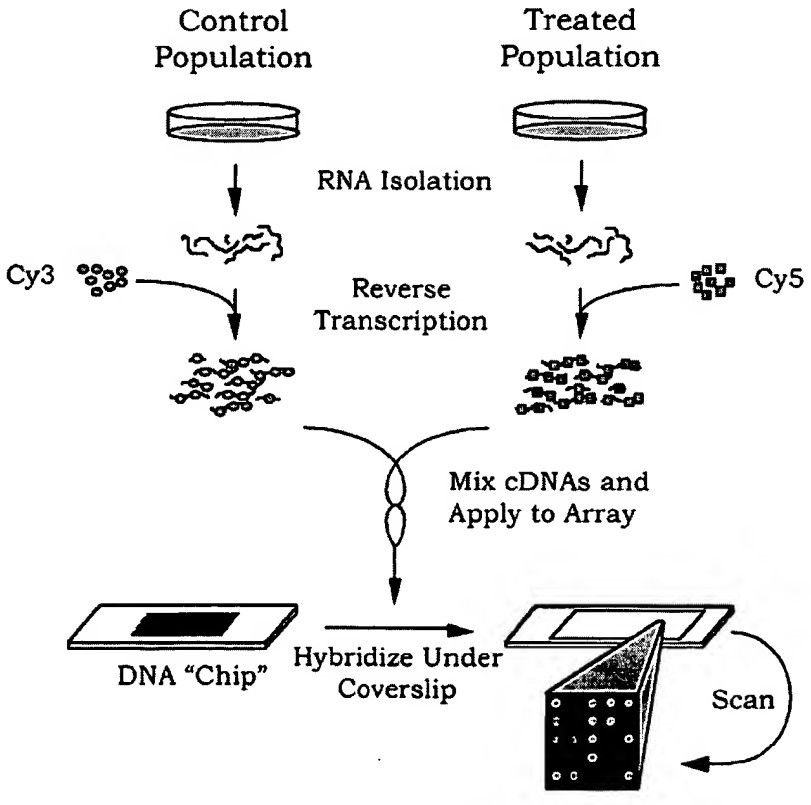


Figure 1. Simplified overview of the method for sample preparation and hybridization to cDNA microarrays. For illustrative purposes, samples derived from cell culture are depicted, although other sample types are amenable to this analysis.

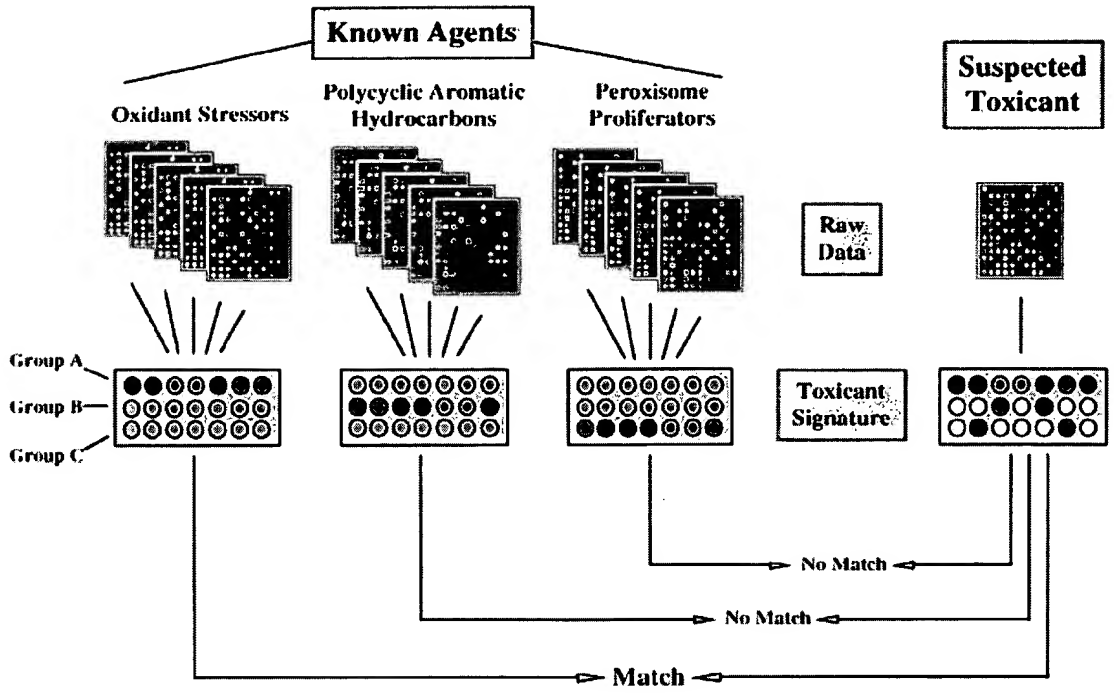


Figure 2. Schematic representation of the method for identification of a toxicant's mechanism of action. In this method, gene-expression data derived from exposure of model systems to known toxicants are analyzed, and a set of changes characteristic to that type of toxicant (termed the toxicant signature) is identified. As depicted, oxidant stressors produce consistent changes in group A genes (indicated by red and green circles), but not group B or C genes (indicated by gray circles). The set of gene-expression changes elicited by the suspected toxicant is then compared with these characteristic patterns, and a putative mechanism of action is assigned to the unknown agent.

The 2090 human genes that comprise this subarray were selected for their well-documented involvement in basic cellular processes as well as their responses to different types of toxic insult. Included on this list are DNA replication and repair genes, apoptosis genes, and genes responsive to PAHs and dioxin-like compounds, peroxisome proliferators, estrogenic compounds, and oxidant stress. Some of the other categories of genes include transcription factors, oncogenes, tumor suppressor genes, cyclins, kinases, phosphatases, cell adhesion and motility genes, and homeobox genes. Also included in this group are 84 housekeeping genes, whose hybridization intensity is averaged and used for signal normalization of the other genes on the chip. To date, very few toxicants have been shown to have appreciable effects on the expression of these housekeeping genes. However, this housekeeping list will be revised if new data warrant the addition or deletion of a particular gene. Table 1 contains a general description of some of the different classes of genes that comprise ToxChip v1.0.

When a toxicant signature is determined, the genes within this signature are flagged within the database. When uncharacterized toxicants are then screened, the data can be quickly reformatted so that blocks of genes representing the different signatures are displayed [11]. This facilitates rapid, visual interpretation of data. We are also developing ToxChip v2.0 and chips for other model systems, including rat, mouse, *Xenopus*, and yeast, for use in toxicology studies.

Animal Models in Toxicology Testing

The toxicology community relies heavily on the use of animals as model systems for toxicology testing. Unfortunately, these assays are inherently expensive, require large numbers of animals and take a long time to complete and analyze. Therefore, the National Institute of Environmental Health Sciences (NIEHS), the National Toxicology Program, and the toxicology community at large are committed to reducing the number of animals used, by developing more efficient and alternative testing methodologies. Although substantial progress has been made in the development of alternative methods, bioassays are still used for testing endpoints such as neurotoxicity, immunotoxicity, reproductive and developmental toxicology, and genetic toxicology. The rodent cancer bioassay is a particularly expensive and time-consuming assay, as it requires almost 4 yr, 1200 animals, and millions of dollars to execute and analyze [43]. In vitro experiments of the type outlined in Figure 2 might provide evidence that an unknown

Table 1. ToxChip v1.0: A Human cDNA Microarray Chip Designed to Detect Responses to Toxic Insult

Gene category	No. of genes on chip
Apoptosis	72
DNA replication and repair	99
Oxidative stress/redox homeostasis	90
Peroxisome proliferator responsive	22
Dioxin/PAH responsive	12
Estrogen responsive	63
Housekeeping	84
Oncogenes and tumor suppressor genes	76
Cell-cycle control	51
Transcription factors	131
Kinases	276
Phosphatases	88
Heat-shock proteins	23
Receptors	349
Cytochrome P450s	30

*This list is intended as a general guide. The gene categories are not unique, and some genes are listed in multiple categories.

agent is (or is not) responsible for eliciting a given biological response. This information would help to select a bioassay more specifically suited to the agent in question or perhaps suggest that a bioassay is not necessary, which would dramatically reduce cost, animal use, and time.

The addition of microarray techniques to standard bioassays may dramatically enhance the sensitivity and interpretability of the bioassay and possibly reduce its cost. Gene-expression signatures could be determined for various types of tissue-specific toxicants, and new compounds could be screened for these characteristic signatures, providing a rapid and sensitive in vivo test. Also, because gene expression is often exquisitely sensitive to low doses of a toxicant, the combination of gene-expression screening and the bioassay might allow the use of lower toxicant doses, which are more relevant to human exposure levels, and the use of fewer animals. In addition, gene-expression changes are normally measured in hours or days, not in the months to years required for tumor development. Furthermore, microarrays might be particularly useful for investigating the relationship between acute and chronic toxicity and identifying secondary effects of a given toxicant by studying the relationship between the duration of exposure to a toxicant and the gene-expression profile produced. Thus, a bioassay that incorporates gene-expression signatures with traditional endpoints might be substantially shorter, use more realistic dose regimens, and cost substantially less than the current assays do.

These considerations are also relevant for branches of toxicology not related to human health and not using rodents as model systems, such as aquatic toxicology and plant pathology. Bioassays based on the flathead minnow, *Daphnia,* and *Arabadopsis* could also be improved by the addition of microarray analysis. The combination of microarrays with traditional bioassays might also be useful for investigating some of the more intractable problems in toxicology research, such as the effects of complex mixtures and the difficulties in cross-species extrapolation.

Exposure Assessment, Environmental Monitoring, and Drug Safety

The currently used methods for assessment of exposure to chemical toxicants are based on measurement of tissue toxin levels or on surrogate markers of toxicity, termed biomarkers (e.g., peripheral blood levels of hepatic enzymes or DNA adducts). Because gene expression is a sensitive endpoint, gene expression as measured with microarray technology may be useful as a new biomarker to more precisely identify hazards and to assess exposure. Similarly, microarrays could be used in an environmental-monitoring capacity to measure the effect of potential contaminants on the gene-expression profiles of resident organisms. In an analogous fashion, microarrays could be used to measure gene-expression endpoints in subjects in clinical trials. The combination of these gene-expression data and more established toxic endpoints in these trials could be used to define highly precise surrogates of safety.

Gene-expression profiles in samples from exposed individuals could be compared to the profiles of the same individuals before exposure. From this information, the nature of the toxic exposure can be determined or a relative clinical safety factor estimated. In the future it may also be possible to estimate not only the nature but the dose of the toxicant for a given exposure, based on relative gene-expression levels. This general approach may be particularly appropriate for occupational-health applications, in which unexposed and exposed samples from the same individuals may be obtainable. For example, a pilot study of gene expression in peripheral-blood lymphocytes of Polish coke-oven workers exposed to PAHs (and many other compounds) is under consideration at the NIEHS. An important consideration for these types of studies is that gene expression can be affected by numerous factors, including diet, health, and personal habits. To reduce the effects of these confounding factors, it may be necessary to compare pools of control samples with pools of treated samples. In the future it may be possible to compare exposed sample sets to a national database of human-expression data, thus eliminating the need to provide an unexposed sample from the same individual. Efforts to develop such a national gene-expression database are currently under way [44,45]. However, this national database approach will require a better understanding of genome-wide gene expression across the highly diverse human population and of the effects of environmental factors on this expression.

Alleles, Oligo Arrays, and Toxicogenetics

Gene sequences vary between individuals, and this variability can be a causative factor in human diseases of environmental origin [46,47]. A new area of toxicology, termed toxicogenetics, was recently developed to study the relationship between genetic variability and toxicant susceptibility. This field is not the subject of this discussion, but it is worthwhile to note that the ability of oligonucleotide arrays to discriminate DNA molecules based on single base-pair differences makes these arrays uniquely useful for this type of analysis. Recent reports demonstrated the feasibility of this approach [41,42]. The NIEHS has initiated the Environmental Genome Project to identify common sequence polymorphisms in 200 genes thought to be involved in environmental diseases [48]. In a pilot study on the feasibility of this application to the Environmental Genome Project, oligonucleotide arrays will be used to resequence 20 candidate genes. This toxicogenetic approach promises to dramatically improve our understanding of interindividual variability in disease susceptibility.

FUTURE PRIORITIES

There are many issues that must be addressed before the full potential of microarrays in toxicology research can be realized. Among these are model system selection, dose selection, and the temporal nature of gene expression. In other words, in which species, at what dose, and at what time do we look for toxicant-induced gene expression? If human samples are analyzed, how variable is global gene expression between individuals, before and after toxicant exposure? What are the effects of age, diet, and other factors on this expression? Experience, in the form of large data sets of toxicant exposures, will answer these questions.

One of the most pressing issues for array scientists is the construction of a national public database (linked to the existing public databases) to serve as a repository for gene-expression data. This relational database must be made available for public use, and researchers must be encouraged to submit their expression data so that others may view and query the information. Researchers at the National Institutes of Health have made laudable progress in developing the first generation of such a database [44,45]. In addition, improved statistical methods for gene clustering and pattern recognition are needed to analyze the data in such a public database.

The proliferation of different platforms and methods for microarray hybridizations will improve sample handling and data collection and analysis and reduce costs. However, the variety of microarray methods available will create problems of data compatibility between platforms. In addition, the near-infinite variety of experimental conditions under which data will be collected by different laboratories will make large-scale data analysis extremely difficult. To help circumvent these future problems, a set of standards to be included on all platforms should be established. These standards would facilitate data entry into the national database and serve as reference points for cross-platform and inter-laboratory data analysis.

Many issues remain to be resolved, but it is clear that new molecular techniques such as microarray hybridization will have a dramatic impact on toxicology research. In the future, the information gathered from microarray-based hybridization experiments will form the basis for an improved method to assess the impact of chemicals on human and environmental health.

ACKNOWLEDGMENTS

The authors would like to thank Drs. Robert Maronpot, George Lucier, Scott Masten, Nigel Walker,Raymond Tennant, and Ms. Theodora Deverenux for critical review of this manuscript. EFN was supported in part by NIEHS Training Grant #ES07017-24.

REFERENCES

1. http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html
2. http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html
3. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269:496–512.
4. Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. Science 1996;274:546, 563–567.
5. http://www.perkin-elmer.com/press/prc5448.html
6. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 1992;257:967–971.
7. Pietu G, Alibert O, Guichard V, et al. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. Genome Res 1996;6:492–503.
8. Zhao ND, Hashida H, Takahashi N, Misumi Y, Sakaki Y. High-density cDNA filter analysis—A novel approach for large-scale, quantitative analysis of gene expression. Gene 1995;156:207–213.
9. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:484–487.
10. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene-expression patterns with a complementary DNA microarray. Science 1995;270:467–470.
11. DeRisi J, Penland L, Brown PO, et al. use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996;14:457–460.
12. Wodicka L, Dong HL, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in Saccharomyces cerevisiae. Nat Biotechnol 1997;15:1359–1367.
13. Marshall A, Hodgson J. DNA chips: An array of possibilities. Nat Biotechnol 1998;16:27–31.
14. http://www.synteni.com
15. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 1996;6:639–645.
16. Chen Y, Dougherty ER, Bittner ML. Ratio-based decisions and the quantitative analysis of cDNA microarray images. Biomedical Optics 1997;2:364–374.
17. Khan J, Simon R, Bittner M, et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 1998;58:5009–5013.
18. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 1996; 93:10614–10619.

19. Lashkari DA, DeRisi JL, McCusker JH, et al. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci USA 1997;94:13057–13062.
20. Heller RA, Schena M, Chai A, et al. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA 1997;94:2150–2155.
21. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997;278:680–686.
22. Drmanac S, Stavropoulos NA, Labat I, et al. Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. Genomics 1996;37:29–40.
23. Milosavljevic A, Savkovic S, Crkvenjakov R, et al. DNA sequence recognition by hybridization to short oligomers: Experimental verification of the method on the E. coli genome. Genomics 1996;37:77–86.
24. Drmanac S, Drmanac R. Processing of cDNA and genomic kilobase-size clones for massive screening, mapping and sequencing by hybridization. Biotechniques 1994;17:328–329, 332–336.
25. http://www.resgen.com/
26. http://www.genomesystems.com/
27. http://www.clontech.com/
28. Pease AC, Solas DA, Fodor SPA. Parallel synthesis of spatially addressable oligonucleotide probe matrices. Abstract. Abstracts of Papers of the American Chemical Society 1992;203:34.
29. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SPA. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 1994;91:5022–5026.
30. Fodor SPA, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991;251:767–773.
31. McGall G, Labadie J, Brock P, Wallraff G, Nguyen T, Hinsberg W. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. Proc Natl Acad Sci USA 1996;93:13555–13560.
32. Lipshutz RJ, Morris D, Chee M, et al. Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 1995;19:442–447.
33. Lockhart DJ, Dong HL, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996;14:1675–1680.
34. http://www.mdyn.com/
35. Sapolsky RJ, Lipshutz RJ. Mapping genomic library clones using oligonucleotide arrays. Genomics 1996;33:445–456.
36. Chee M, Yang R, Hubbell E, et al. Accessing genetic information with high-density DNA arrays. Science 1996;274:610–614.
37. Hacia JG, Makalowski W, Edgemon K, et al. Evolutionary sequence comparisons using high-density oligonucleotide arrays. Nat Genet 1998;18:155–158.
38. Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM, Miyada CG. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum Mutat 1996;7:244–255.
39. Hacia JG, Brody LC, Chee MS, Fodor SPA, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nat Genet 1996;14:441–447.
40. Kozal MJ, Shah N, Shen NP, et al. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat Med 1996;2:753–759.
41. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998;280:1077–1082.
42. Winzeler EA, Richards DR, Conway AR, et al. Direct allelic variation scanning of the yeast genome. Science 1998;281:1194–1197.
43. Chhabra RS, Huff JE, Schwetz BS, Selkirk J. An overview of prechronic and chronic toxicity carcinogenicity experimental-study designs and criteria used by the National Toxicology Program. Environ Health Perspect 1990;86:313–321.
44. Ermolaeva O, Rastogi M, Pruitt KD, et al. Data management and analysis for gene expression arrays. Nat Genet 1998;20:19–23.
45. http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/dbase.html
46. Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 1996;382:722–725.
47. Bell DA, Taylor JA, Paulson DF, Robertson CN, Mohler JL, Lucier GW. Genetic risk and carcinogen exposure—A common inherited defect of the carcinogen-metabolism gene glutathione-S-transferase M1 (Gstm1) that increases susceptibility to bladder cancer. J Natl Cancer Inst 1993;85:1159–1164.
48. http://www.niehs.nih.gov/envgenom/home.html

ELSEVIER

Toxicology Letters 112–113 (2000) 467–471

Toxic l gy Letters

www.elsevier.com/locate/toxlet

Expression profiling in toxicology — potentials and limitations

Sandra Steiner *, N. Leigh Anderson

Large Scale Biology Corporation, 9620 Medical Center Drive, Rockville, MD 20850-3338, USA

Abstract

Recent progress in genomics and proteomics technologies has created a unique opportunity to significantly impact the pharmaceutical drug development processes. The perception that cells and whole organisms express specific inducible responses to stimuli such as drug treatment implies that unique expression patterns, molecular fingerprints, indicative of a drug's efficacy and potential toxicity are accessible. The integration into state-of-the-art toxicology of assays allowing one to profile treatment-related changes in gene expression patterns promises new insights into mechanisms of drug action and toxicity. The benefits will be improved lead selection, and optimized monitoring of drug efficacy and safety in pre-clinical and clinical studies based on biologically relevant tissue and surrogate markers.

Keywords: Proteomics; Genomics; Toxicology

1. Introduction

The majority of drugs act by binding to protein targets, most to known proteins representing enzymes, receptors and channels, resulting in effects such as enzyme inhibition and impairment of signal transduction. The treatment-induced perturbations provoke feedback reactions aiming to compensate for the stimulus, which almost always are associated with signals to the nucleus, resulting in altered gene expression. Such gene expression regulations account for both the pharmacological action and the toxicity of a drug and can be visualized by either global mRNA or global protein expression profiling. Hence, for each individual drug, a characteristic gene regulation pattern, its molecular fingerprint, exists which bears valuable information on its mode of action and its mechanism of toxicity.

Gene expression is a multistep process that results in an active protein (Fig. 1). There exist numerous regulation systems that exert control at and after the transcription and the translation step. Genomics, by definition, encompasses the quantitative analysis of transcripts at the mRNA level, while the aim of proteomics is to quantify gene expression further down-stream, creating a snapshot of gene regulation closer to ultimate cell function control.

* Corresponding author. Tel.: +1-301-4245989; fax: +1-301-7624892.
E-mail address: steiner@lsbc.com (S. Steiner)

0378-4274/00/$ - see front matter
PII: S0378-4274(99)00236-2

2. Global mRNA profiling

Expression data at the mRNA level can be produced using a set of different technologies such as DNA microarrays, reverse transcript imaging, amplified fragment length polymorphism (AFLP), serial analysis of gene expression (SAGE) and others. Currently, DNA microarrays are very popular and promise a great potential. On a typical array, each gene of interest is represented either by a long DNA fragment (200–2400 bp) typically generated by polymerase chain reaction (PCR) and spotted on a suitable substrate using robotics (Schena et al., 1995; Shalon et al., 1996) or by several short oligonucleotides (20–30 bp) synthesized directly onto a solid support using photolabile nucleotide chemistry (Fodor et al., 1991; Chee et al., 1996). From control and treated tissues, total RNA or mRNA is isolated and reverse transcribed in the presence of radioactive or fluorescent labeled nucleotides, and the labeled probes are then hybridized to the arrays. The intensity of the array signal is measured for each gene transcript by either autoradiography or laser scanning confocal microscopy. The ratio between the signals of control and treated samples reflect the relative drug-induced change in transcript abundance.

3. Global protein profiling

Global quantitative expression analysis at the protein level is currently restricted to the use of two-dimensional gel electrophoresis. This technique combines separation of tissue proteins by isoelectric focusing in the first dimension and by sodium dodecyl sulfate slab gel electrophoresis-based molecular weight separation on the second, orthogonal dimension (Anderson et al., 1991). The product is a rectangular pattern of protein spots that are typically revealed by Coomassie Blue, silver or fluorescent staining (Fig. 2). Protein spots are identified by mass spectrometry following generation of peptide mass fingerprints (Mann et al., 1993) and sequence tags (Wilkins et al., 1996). Similar to the mRNA approach, the ratio between the optical density of spots from control and treated samples are compared to search for treatment-related changes.

4. Expression data analysis

Bioinformatics forms a key element required to organize, analyze and store expression data from either source, the mRNA or the protein level. The overall objective, once a mass of high-quality

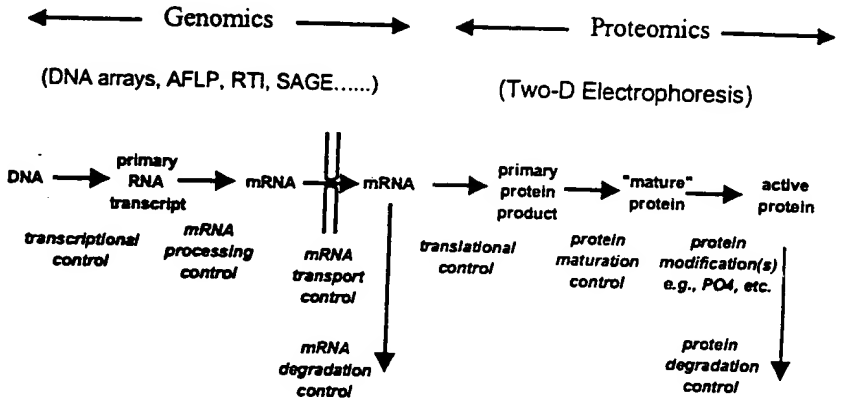


Fig. 1. Production of an active protein is a multistep process in which numerous regulation systems exert control at various stages of expression. Molecular fingerprints of drugs can be visualized through expression profiling at the mRNA level (genomics) using a variety of technologies and at the protein level (proteomics) using two-dimensional gel electrophoresis.

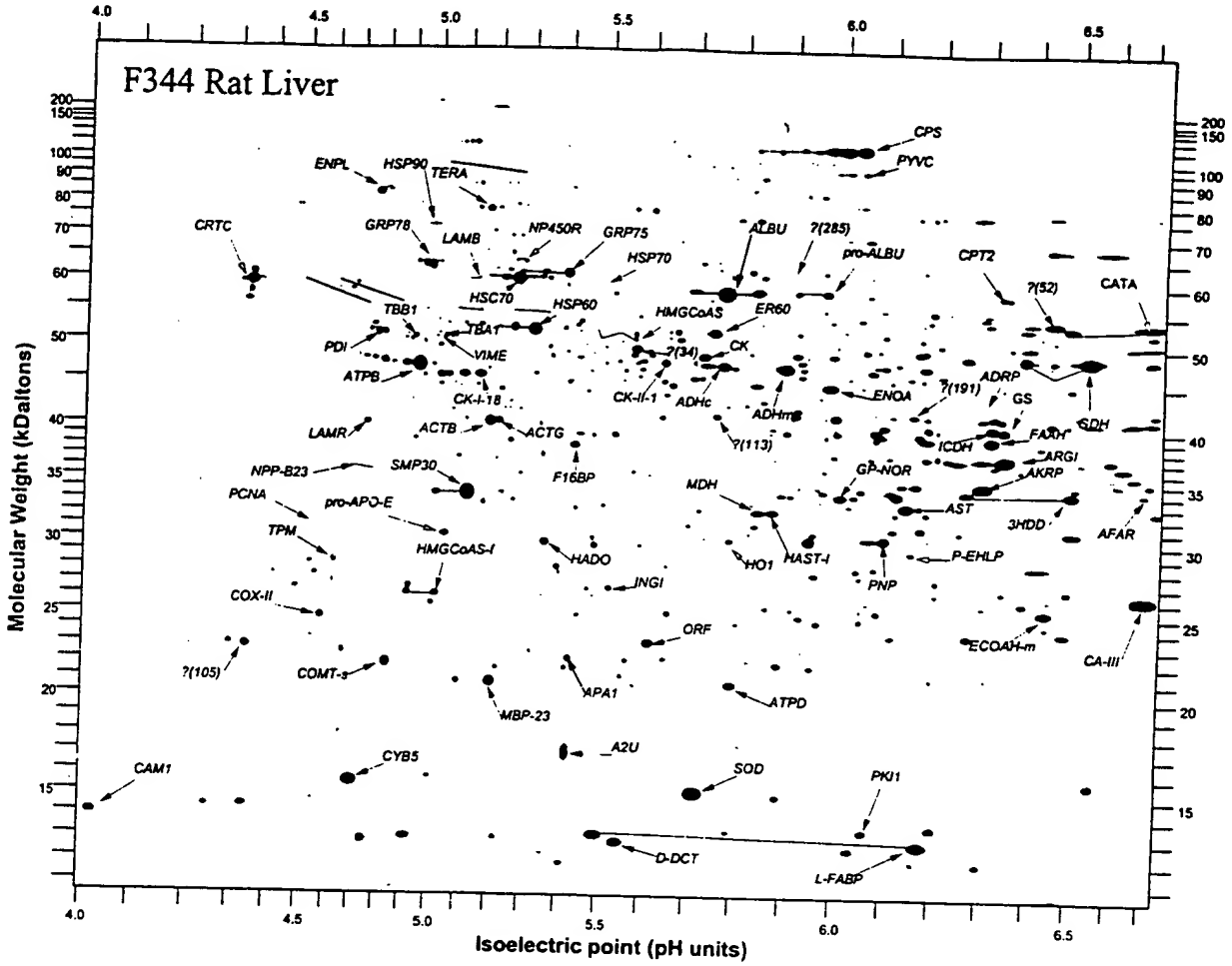


Fig. 2. Computerized representation of a Coomassie Blue stained two-dimensional gel electrophoresis pattern of Fischer F344 rat liver homogenate.

quantitative expression data has been collected, is to visualize complex patterns of gene expression changes, to detect pathways and sets of genes tightly correlated with treatment efficacy and toxicity, and to compare the effects of different sets of treatment (Anderson et al., 1996). As the drug effect database is growing, one may detect similarities and differences between the molecular fingerprints produced by various drugs, information that may be crucial to make a decision whether to refocus or extend the therapeutic spectrum of a drug candidate.

5. Comparison of global mRNA and protein expression profiling

There are several synergies and overlaps of data obtained by mRNA and protein expression analysis. Low abundant transcripts may not be easily quantified at the protein level using standard two-dimensional gel electrophoresis analysis and their detection may require prefractionation of samples. The expression of such genes may be preferably quantified at the mRNA level using techniques allowing PCR-mediated target amplifi-

cation. Tissue biopsy samples typically yield good quality of both mRNA and proteins; however, the quality of mRNA isolated from body fluids is often poor due to the faster degradation of mRNA when compared with proteins. RNA samples from body fluids such as serum or urine are often not very 'meaningful', and secreted proteins are likely more reliable surrogate markers for treatment efficacy and safety. Detection of post-translational modifications, events often related to function or nonfunction of a protein, is restricted to protein expression analysis and rarely can be predicted by mRNA profiling. Information on subcellular localization and translocation of proteins has to be acquired at the level of the protein in combination with sample prefractionation procedures. The growing evidence of a poor correlation between mRNA and protein abundance (Anderson and Seilhamer, 1997) further suggests that the two approaches, mRNA and protein profiling, are complementary and should be applied in parallel.

6. Expression profiling and drug development

Understanding the mechanisms of action and toxicity, and being able to monitor treatment efficacy and safety during trials is crucial for the successful development of a drug. Mechanistic insights are essential for the interpretation of drug effects and enhance the chances of recognizing potential species specificities contributing to an improved risk profile in humans (Richardson et al., 1993; Steiner et al., 1996b; Aicher et al., 1998). The value of expression profiling further increases when links between treatment-induced expression profiles and specific pharmacological and toxic endpoints are established (Anderson et al., 1991, 1995, 1996; Steiner et al. 1996a). Changes in gene expression are known to precede the manifestation of morphological alterations, giving expression profiling a great potential for early compound screening, enabling one to select drug candidates with wide therapeutic windows reflected by molecular fingerprints indicative of high pharmacological potency and low toxicity (Arce et al., 1998). In later phases of drug development, surrogate markers of treatment efficacy and toxicity can be applied to optimize the monitoring of pre-clinical and clinical studies (Doherty et al., 1998).

7. Perspectives

The basic methodology of safety evaluation has changed little during the past decades. Toxicity in laboratory animals has been evaluated primarily by using hematological, clinical chemistry and histological parameters as indicators of organ damage. The rapid progress in genomics and proteomics technologies creates a unique opportunity to dramatically improve the predictive power of safety assessment and to accelerate the drug development process. Application of gene and protein expression profiling promises to improve lead selection, resulting in the development of drug candidates with higher efficacy and lower toxicity. The identification of biologically relevant surrogate markers correlated with treatment efficacy and safety bears a great potential to optimize the monitoring of pre-clinical and clinical trails.

References

Aicher, L., Wahl, D., Arce, A., Grenet, O., Steiner, S., 1998. New insights into cyclosporine A nephrotoxicity by proteome analysis. Electrophoresis 19, 1998–2003.

Anderson, N.L., Seilhamer, J., 1997. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 18, 533–537.

Anderson, N.L., Esquer-Blasco, R., Hofmann, J.P., Anderson, N.G., 1991. A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. Electrophoresis 12, 907–930.

Anderson, L., Steele, V.K., Kelloff, G.J., Sharma, S., 1995. Effects of oltipraz and related chemoprevention compounds on gene expression in rat liver. J. Cell. Biochem. Suppl. 22, 108–116.

Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., 1996. The effects of peroxisome proliferators on protein abundances in mouse liver. Toxicol. Appl. Pharmacol. 137, 75–89.

Arce, A., Aicher, L., Wahl, D., Esquer-Blasco, R., Anderson, N.L., Cordier, A., Steiner, S., 1998. Changes in the liver proteome of female Wistar rats treated with the hypoglycemic agent SDZ PGU 693. Life Sci. 63, 2243–2250.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., Fodor, S.P., 1996. Accessing genetic information with high-density DNA arrays. Science 274, 610–614.

Doherty, N.S., Littman, B.H., Reilly, K., Swindell, A.C., Buss, J., Anderson, N.L., 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis 19, 355–363.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D., 1991. Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767–773.

Mann, M., Hojrup, P., Roepsdorff, P., 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. 22, 338–345.

Richardson, F.C., Strom, S.C., Copple, D.M., Bendele, R.A., Probst, G.S., Anderson, N.L., 1993. Comparisons of protein changes in human and rodent hepatocytes induced by the rat-specific carcinogen, methapyrilene. Electrophoresis 14, 157–161.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expresssion patterns with a complementary DNA microarray. Science 251, 467–470.

Shalon, D., Smith, S.J., Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6, 639–645.

Steiner, S., Wahl, D., Mangold, B.L.K., Robison, R., Rayrnackers, J., Meheus, L., Anderson, N.L., Cordier, A., 1996a. Induction of the adipose differentiation-related protein in liver of etomoxir treated rats. Biochem. Biophys. Res. Commun. 218, 777–782.

Steiner, S., Aicher, L., Raymackers, J., Meheus, L., Esquer-Blasco, R., Anderson, L., Cordier, A., 1996b. Cyclosporine A mediated decrease in the rat renal calcium binding protein calbindin-D 28 kDa. Biochem. Pharmacol. 51, 253–258.

Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Appel, R.D., Hochstrasser, D.F., 1996. Protein identification with sequence tags. Curr. Biol. 6, 1543–1544.

Docket No.: PF-0636 RCE
USSN: 09/831,458
Ref. No. 5 of 17

Application of DNA Arrays to Toxicology

John C. Rockett and David J. Dix

Reproductive Toxicology Division, National Health and Environmental Effects Research Laborat ry, U.S. Envir nmental Protection Agency, Research Triangle Park, North Carolina, USA

DNA array technology makes it possible to rapidly genotype individuals or quantify the expression of thousands of genes on a single filter or glass slide, and holds enormous potential in toxicologic applications. This potential led to a U.S. Environmental Protection Agency-sponsored workshop titled "Application of Microarrays to Toxicology" on 7–8 January 1999 in Research Triangle Park, North Carolina. In addition to providing state-of-the-art information on the application of DNA or gene microarrays, the workshop catalyzed the formation of several collaborations, committees, and user's groups throughout the Research Triangle Park area and beyond. Potential application of microarrays to toxicologic research and risk assessment include genome-wide expression analyses to identify gene-expression networks and toxicant-specific signatures that can be used to define mode of action, for exposure assessment, and for environmental monitoring. Arrays may also prove useful for monitoring genetic variability and its relationship to toxicant susceptibility in human populations. *Key words:* DNA arrays, gene arrays, microarrays, toxicology. *Environ Health Perspect* 107:681–685 (1999). [Online 6 July 1999]
http://ehpnet1.niehs.nih.gov/docs/1999/107p681-685rockett/abstract.html

Decoding the genetic blueprint is a dream that offers manifold returns in terms of understanding how organisms develop and function in an often hostile environment. With the rapid advances in molecular biology over the last 30 years, the dream has come a step closer to reality. Molecular biologists now have the ability to elucidate the composition of any genome. Indeed, almost 20 genomes have already been sequenced and more than 60 are currently under way. Foremost among these is the Human Genome Mapping Project. However, the genomes of a number of commonly used laboratory species are also under intensive investigation, including yeast, *Arabidopsis*, maize, rice, zebra fish, mouse, rat, and dog. It is widely expected that the completion of such programs will facilitate the development of many powerful new techniques and approaches to diagnosing and treating genetically and environmentally induced diseases which afflict mankind. However, the vast amount of data being generated by genome mapping will require new high-throughput technologies to investigate the function of the millions of new genes that are being reported. Among the most widely heralded of the new functional genomics technologies are DNA arrays, which represent perhaps the most anticipated new molecular biology technique since polymerase chain reaction (PCR).

Arrays enable the study of literally thousands of genes in a single experiment. The potential importance of arrays is enormous and has been highlighted by the recent publication of an entire *Nature Genetics* supplement dedicated to the technology (*1*). Despite this huge surge of interest, DNA arrays are still little used and largely unproven, as demonstrated by the high ratio of review and press articles to actual data papers. Even so, the potential they offer has driven venture capitalists into a frenzy of investment and many new companies are springing up to claim a share of this rapidly developing market.

The U.S. Environmental Protection Agency (EPA) is interested in applying DNA array technology to ongoing toxicologic studies. To learn more about the current state of the technology, the Reproductive Toxicology Division (RTD) of the National Health and Environmental Effects Research Laboratory (NHEERL; Research Triangle Park, NC) hosted a workshop on "Application of Microarrays to Toxicology" on 7–8 January 1999 in Research Triangle Park, North Carolina. The workshop was organized by David Dix, Robert Kavlock, and John Rockett of the RTD/NHEERL. Twenty-two intramural and extramural scientists from government, academia, and industry shared information, data, and opinions on the current and future applications for this exciting new technology. The workshop had more than 150 attendees, including researchers, students, and administrators from the EPA, the National Institute of Environmental Health Sciences (NIEHS), and a number of other establishments from Research Triangle Park and beyond. Presentations ranged from the technology behind array production through the sharing of actual experimental data and projections on the future importance and applications of arrays. The information contained in the workshop presentations should provide aid and insight into arrays in general and their application to toxicology in particular.

Array Elem nts

In the context of molecular biology, the word "array" is normally used to refer to a series of DNA or protein elements firmly attached in a regular pattern to some kind of supportive medium. DNA array is often used interchangeably with gene array or microarray. Although not formally defined, microarray is generally used to describe the higher density arrays typically printed on glass chips. The DNA elements that make up DNA arrays can be oligonucleotides, partial gene sequences, or full-length cDNAs. Companies offering pre-made arrays that contain less than full-length clones normally use regions of the genes which are specific to that gene to prevent false positives arising through cross-hybridization. Sequence verification of cDNA clone identity is necessary because of errors in identifying specific clones from cDNA libraries and databases. Premade DNA arrays printed on membranes are currently or imminently available for human, mouse, and rat. In most cases they contain DNA sequences representing several thousand different sequence clusters or genes as delineated through the National Center for Biotechnology Information UniGene Project (*2*). Many of these different UniGene clusters (putative genes) are represented only by expressed sequence tags (ESTs).

Array Printing

Arrays are typically printed on one of two types of support matrix. Nylon membranes are used by most off-the-shelf array providers such as Clontech Laboratories, Inc. (Palo Alto, CA), Genome Systems, Inc. (St. Louis, MO), and Research Genetics, Inc. (Huntsville, AL). Microarrays such as those produced by Affymetrix, Inc. (Santa Clara, CA), Incyte Pharmaceuticals, Inc. (Palo Alto, CA), and many do-it-yourself (DIY) arraying groups use glass wafers or slides. Although standard microscope slides may be used, they must be preprepared to facilitate sticking of the DNA to the glass. Several different

Address correspondence to J. Rockett, Reproductive Toxicology Division (MD-72), National Health and Environmental Effects Research Laboratory, U.S. EPA, Research Triangle Park, NC 27711 USA. Telephone: (919) 541-2678. Fax: (919) 541-4017. E-mail: rockett.john@epa.gov

The authors thank R. Kavlock for envisioning the application of array technology to toxicology at the U.S. Environmental Protection Agency. We also thank T. Wall and B. Deitz for administrative assistance.

This document has been reviewed in accordance with EPA policy and approved for publication. Mention of companies, trade names, or products does not signify endorsement of such by the EPA.

Received 23 March 1999; accepted 22 April 1999.

coatings have been successfully used, including silane and lysine. The coating of slides can easily be carried out in the laboratory, but many prefer the convenience of precoated slides available from suppliers.

Once the support matrix has been prepared, the DNA elements can be applied by several methods. Affymetrix, Inc., has developed a unique photolithographic technology for attaching oligonucleotides to glass wafers. More commonly, DNA is applied by either noncontact or contact printing. Noncontact printers can use thermal, solenoid, or piezoelectric technology to spray aliquots of solution onto the support matrix and may be used to produce slide or membrane-based arrays. Cartesian Technologies, Inc. (Irvine, CA) has developed nQUAD technology for use in its PixSys printers. The system couples a syringe pump with the microsolenoid valve, a combination that provides rapid quantitative dispensing of nanoliter volumes (down to 4.2 nL) over a variable volume range. A different approach to noncontact printing uses a solid pin and ring combination (Genetic MicroSystems, Inc., Woburn, MA). This system (Figure 1) allows a broader range of sample, including cell suspensions and particulates, because the printing head cannot be blocked up in the same way as a spray nozzle. Fluid transfer is controlled in this system primarily by the pin dimensions and the force of deposition, although the nature of the support matrix and the sample will also affect transfer to some degree.

In contact printing, the pin head is dipped in the sample and then touched to the support matrix to deposit a small aliquot. Split pins were one of the first contact-printing devices to be reported and are the suggested format for DIY arrayers, as described by Brown (*3*). Split pins are small metal pins with a precise groove cut vertically in the middle of the pin tip. In this system, 1–48 split pins are positioned in the pin-head. The split pins work by simple capillary action, not unlike a fountain pen—when the pin heads are dipped in the sample, liquid is drawn into the pin groove. A small (fixed) volume is then deposited each time the split pins are gently touched to the support matrix. Sample (100–500 pL depending on a variety of parameters) can be deposited on multiple slides before refilling is required, and array densities of > 2,500 spots/cm^2 may be produced. The deposit volume depends on the split size, sample fluidity, and the speed of printing. Split pins are relatively simple to produce and can be made in-house if a suitable machine shop is available. Alternatively, they can be obtained directly from companies such as TeleChem International, Inc. (Sunnyvale, CA).

Irrespective of their source, printers should be run through a preprint sequence prior to producing the actual experimental arrays; the first 100 or so spots of a new run tend to be somewhat variable. Factors effecting spot reproducibility include slide treatment homogeneity, sample differences, and instrument errors. Other factors that come into play include clean ejection of the drop and clogging (nQUAD printing) and mechanical variations and long-term alteration in print-head surface of solid and split pins. However, with careful preparation it is possible to get a coefficient of variance for spot reproducibility below 10%.

One potential printing problem is sample carryover. Repeated washing, blotting, and drying (vacuum) of print pins between samples is normally effective at reducing sample carryover to negligible amounts. Printing should also be carried out in a controlled environment. Humidified chambers are available in which to place printers. These help prevent dust contamination and produce a uniform drying rate, which is important in determining spot size, quality, and reproducibility.

In summary, although several printing technologies are available, none are particularly outstanding and the bottom line is that they are still in a relatively early stage of evolution.

Array Hybridization

The hybridization protocol is, practically speaking, relatively straightforward and those with previous experience in blotting should have little difficulty. Array hybridizations are, in essence, reverse Southern/Northern blots—instead of applying a labeled probe to the target population of DNA/RNA, the labeled population is applied to the probe(s). With membrane-based arrays, the control and treated mRNA populations are normally converted to cDNA and labeled with isotope (e.g., ^{33}P) in the process. These labeled populations are then hybridized independently to parallel or serial arrays and the hybridization signal is detected with a phosporimager. A less commonly used alternative to radioactive probes is enzymatic detection. The probe may be biotinylated, haptenylated, or have alkaline phosphatase/horseradish peroxidase attached. Hybridization is detected by enzymatic reaction yielding a color reaction (*4*). Differences in hybridization signals can be detected by eye or, more accurately, with the help of digital imaging and commercially available software. The labeling of the test populations for slide-based microarrays uses a slightly different approach. The probe typically consists of two samples of polyA$^+$ RNA (usually from a treated and a control population) that are converted to cDNA; in the process each is labeled with a different fluor. The independently labeled probes are then mixed together and hybridized to a single microarray slide and the resulting combined fluorescent signal is scanned. After normalization, it is possible to determine the ratio of fluorescent signals from a single hybridization of a slide-based microarray.

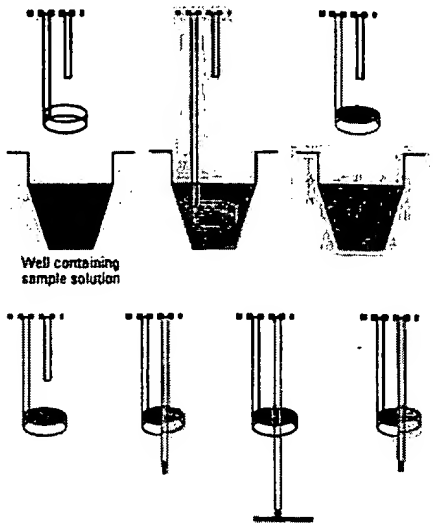


Figure 1. Genetic Microsystems (Woburn, MA) pin ring system for printing arrays. The pin ring combination consists of a circular open ring oriented parallel to the sample solution, with a vertical pin centered over the ring. When the ring is dipped into a solution and lifted, it withdraws an aliquot of sample held by surface tension. To spot the sample, the pin is driven down through the ring and a portion of the solution is transferred to the bottom of the pin. The pin continues to move downward until the pendant drop of solution makes contact with the underlying surface. The pin is then lifted, and gravity and surface tension cause deposition of the spot onto the array. Figure from Flowers et al. (*14*), with permission from Genetic Microsystems.

cDNA derived from control and treated populations of RNA is most commonly hybridized to arrays, although subtractive hybridization or differential display reactions may also be used. Fluorophore- or radiolabeled nucleotides are directly incorporated into the cDNA in the process of converting RNA to cDNA. Alternatively, 5′ end-labeled primers may be used for cDNA synthesis. These are labeled with a fluorophore for direct visualization of the hybridized array. Alternatively, biotin or a hapten may be attached to the primer, in which case fluor-labeled streptavidin or antibody must be applied before a signal can be generated. The most commonly used fluorophores at present are cyanine (Cy)3 and Cy5 (Amersham Pharmacia Biotech AB, Uppsala, Sweden). However, the relative expense of these fluorescent conjugates has driven a search for cheaper alternatives. Fluorescein, rhodamine, and Texas red have all been used, and companies such as Molecular Probes, Inc. (Eugene, OR) are developing a series of labeled nucleotides with a wide range of excitation and emission spectra which may prove to function as well as the Cy dyes.

Analysis of DNA Microarrays

Membrane-based arrays are normally analyzed on film or with a phosphorimager, whereas chip-based arrays require more specialized scanning devices. These can be divided into three main groups: the charge-coupled device camera systems, the nonconfocal laser scanners, and the confocal laser scanners. The advantages and disadvantages of each system are listed in Table 1.

Because a typical spot on a microarray can contain > 10^8 molecules, it is clear that a large variation in signal strength may occur. Current scanners cannot work across this many orders of magnitude (4 or 5 is more typical). However, the scanning parameters can normally be adjusted to collect more or less signal, such that two or three scans of the same array should permit the detection of rare and abundant genes.

When a microarray is scanned, the fluorescent images are captured by software normally included with the scanner. Several commercial suppliers provide additional software for quantifying array images, but the software tools are constantly evolving to meet the developing needs of researchers, and it is prudent to define one's own needs and clarify the exact capabilities of the software before its purchase. Issues that should be considered include the following:

- Can the software locate offset spots?
- Can it quantitate across irregular hybridization signals?
- Can the arrayed genes be programmed in for easy identification and location?
- Can the software connect via the Internet to databases containing further information on the gene(s) of interest?

One of the key issues raised at the workshop was the sensitivity of microarray technology. Experiments by General Scanning, Inc. (Watertown, MA), have shown that by using the Cy dyes and their scanner, signal can be detected down to levels of < 1 fluor molecule per square micrometer, which translates to detecting a rare message at approximately one copy per cell or less.

Array Applications

Although arrays are an emerging technology certain to undergo improvement and alteration, they have already been applied usefully to a number of model systems. Arrays are at their most powerful when they contain the entire genome of the species they are being used to study. For this reason, they have strong support among researchers utilizing yeast and *Caenorhabditis elegans* (*5*). The genomes of both of these species have been sequenced and, in the case of yeast, deposited onto arrays for examination of gene expression (*6,7*). With both of these species, it is relatively easy to perturb individual gene expression. Indeed, *C. elegans* knockouts can be made simply by soaking the worms in an antisense solution of the gene to be knocked out.

Table 1. Advantages and disadvantages of different microarray scanning systems.

		Nonconfocal laser scanner	
Advantages	Few moving parts	Relatively simple optics	Small depth of focus reduces artifacts
	Fast scanning of bright samples		May have high light collection efficiency
Disadvantages	Less appropriate for dim samples	Low light collection efficiency	Small depth of focus requires scanning precision
	Optical scatter can limit performance	Background artifacts not rejected	
		Resolution typically low	

CCD, charge-coupled device.
From Kawasaki (*13*).

By a process of systematic gene disruption, it is now possible to examine the cause and effect relationships between different genes in these simple organisms. This kind of approach should help elucidate biochemical pathways and genetic control processes, deconvolute polygenic interactions, and define the architecture of the cellular network. A simple case study of how this can be achieved was presented by Butow [University of Texas Southwestern Medical Center, Dallas, TX (Figure 2)]. Although it is the phenotypic result of a single gene knockout that is being examined, the effect of such perturbation will almost always be polygenic. Polygenic interactions will become increasingly important as researchers begin to move away from single gene systems when examining the nature of toxicologic responses to external stimuli. This is especially important in toxicology because the phenotype produced by a given environmental insult is never the result of the action of a single gene; rather, it is a complex interaction of one or multiple cellular pathways. Phenomena such as quantitative trait (the continuous variation of phenotype), epistasis (the effect of alleles of one or more genes on the expression of other genes), and penetrance (proportion of individuals of a given genotype that display a particular phenotype) will become increasingly evident and important as toxicologists push toward the ultimate goal of matching the responses of individuals to different environmental stimuli.

Analysis of the transcriptome (the expression level of all the genes in a given cell population) was a use of arrays addressed by several speakers. Unfortunately, current gene nomenclature is often confusing in that single genes are allocated multiple names (usually as a result of independent discovery by different laboratories), and there was a call for standardization of gene nomenclature. Nevertheless, once a transcriptome has been assembled it can then be transferred onto arrays and used to screen any chosen system. The EPA MicroArray Consortium (EPAMAC) is assembling testes transcriptomes for human, rat, and mouse. In a slightly different approach, Nuwaysir et al. (*8*) describes how the NIEHS assembled what is effectively a "toxicological transcriptome"—a library of human and mouse genes that have previously been proven or implicated in responses to toxicologic insults. Clontech Laboratories, Inc. (Palo Alto, CA), has begun a similar process by developing stress/toxicology filter arrays of rat, mouse, and human genes. Thus, rather than being tissue or cell specific, these stress/toxicology arrays can be used across a variety of model systems to look for alterations in the expression of toxicologically important genes and define the new field of toxicogenomics. The potential to identify toxicant families based on tissue- or cell-specific gene expression could revolutionize drug testing. These molecular signatures or fingerprints could not only point to the possible toxicity/carcinogenicity of newly discovered compounds (Figure 3), but also aid in elucidating their mechanism of action through identification of gene expression networks. By extension, such signatures could provide easily identifiable biomarkers to assess the degree, time, and nature of exposure.

DNA arrays are primarily a tool for examining differential gene expression in a given model. In this context they are referred to as closed systems because they lack the ability of other differential expression technologies, e.g., differential display and subtractive hybridization, to detect previously unknown genes not present on the array. This would appear to limit the power of DNA arrays to the imaginations and preconceptions of the researcher in selecting genes previously characterized and thought to be involved in the model system. However, the various genome sequencing projects have created a new category of sequence—the EST—that has partially mollified this deficiency. ESTs are cDNAs expressed in a given tissue that, although they may share some degree of sequence similarity to previously characterized genes, have not been assigned specific genetic identity. By incorporating EST clones into an array, it is possible to monitor the expression of these unknown genes. This can enable the identification of previously uncharacterized genes that may have biologic

significance in the model system. Filter arrays from Research Genetics and slide arrays from Incyte Pharmaceuticals both incorporate large numbers of ESTs from a variety of species.

A further use of microarrays is the identification of single nucleotide polymorphisms (SNPs). These genomic variations are abundant—they occur approximately every 1 kb or so—and are the basis of restriction fragment length polymorphism analysis used in forensic analysis. Affymetrix, Inc., designed chips that contain multiple repeats of the same gene sequence. Each position is present with all four possible bases. After the hybridization of the sample, the degree of hybridization to the different sequences can be measured and the exact sequence of the target gene deduced. SNPs are thought to be of vital importance in drug metabolism and toxicology. For example, single base differences in the regulatory region or active site of some genes can account for huge differences in the activity of that gene. Such SNPs are thought to explain why some people are able to metabolize certain xenobiotics better than others. Thus, arrays provide a further tool for the toxicologist investigating the nature of susceptible subpopulations and toxicologic response.

There are still many wrinkles to be ironed out before arrays become a standard tool for toxicologists. The main issues raised at the workshop by those with hands-on experience were the following:

- Expense: the cost of purchasing/contracting this technology is still too great for many individual laboratories.
- Clones: the logistics of identifying, obtaining, and maintaining a set of nonredundant, noncontaminated, sequence-verified, species/cell/tissue/field-specific clones.
- Use of inbred strains: where whole-organism models are being used, the use of inbred strains is important to reduce the potentially confusing effects of the individual variation typically seen in outbred populations.
- Probe: the need for relatively large amounts of RNA, which limits the type of sample (e.g., biopsy) that can be used. Also, different RNA extraction methods can give different results.
- Specificity: the ability to discriminate accurately between closely related genes (e.g., the cytochrome p450 family) and splice variants.
- Quantitation: the quantitation of gene expression using gene arrays is still open to debate. One reason for this is the different incorporation of the labeling dyes. However, the main difficulty lies in knowing what to normalize against. One option is to include a large number of so-called housekeeping genes in the array. However, the expression of these genes often change depending on the tissue and the toxicant, so it is necessary to characterize the expression of these genes in the model system before utilizing them. This is clearly not a viable option when screening multiple new compounds. A second option is to include on the array genes from a nonrelated species (e.g., a plant gene on an animal array) and to spike the probe with synthetic RNA(s) complementary to the gene(s).
- Reproducibility: this is sometimes questionable, and a figure of approximately two or three repeats was used as the minimum number required to confirm initial findings. Again, however, most people advocated the use of Northern blots or reverse transcriptase PCR to confirm findings.
- Sensitivity: concerns were voiced about the number of target molecules that must be present in a sample for them to be detected on the array.
- Efficiency: reproducible identification of 1.5- to 2-fold differences in expression was reported, although the number of genes that undergo this level of change and remain undetected is open to debate. It is important that this level of detection be ultimately achieved because it is commonly perceived that some important transcription factors and their regulators respond at such low levels. In most cases, 3- to 5-fold was the minimum change that most were happy to accept.
- Bioinformatics: perhaps the greatest concern was how to accurately interpret the data with the greatest accuracy and efficiency. The biggest headache is trying to identify networks of gene expression that are common to different treatments or doses. The amount of data from a single experiment is huge. It may be that, in the future, several groups individually equipped with specialized software algorithms for studying their favorite genes or gene systems will be able to share the same hybridized chips. Thus, arrays could usher in a new perspective on collaboration and the sharing of data.

EPAMAC

Perhaps the main reason most scientists are unable to use array technology is the high cost involved, whether buying off-the-shelf membranes, using contract printing services, or

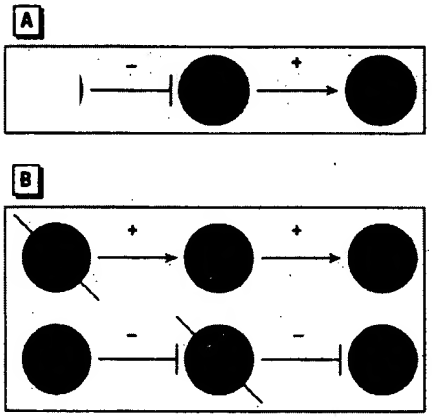


Figure 2. Potential effects of gene knockout within positively and negatively regulated gene expression networks. i_1 is limiting in wild type for expression of i_2. (*A*) A simple, two-component, linear regulatory network operating on gene i_2, where i_1 is a positive effector of i_2 and i_n is either a positive or negative effector of i_1. This network could be deduced by examining the consequence of (*B*) deleting i_n on the expression of i_1 and i_2, where the expression of i_2 would be decreased or increased depending on whether i_n was a positive or negative regulator. These and other connected components of even greater complexity could be revealed by genome-wide expression analysis. From Butow (*15*).

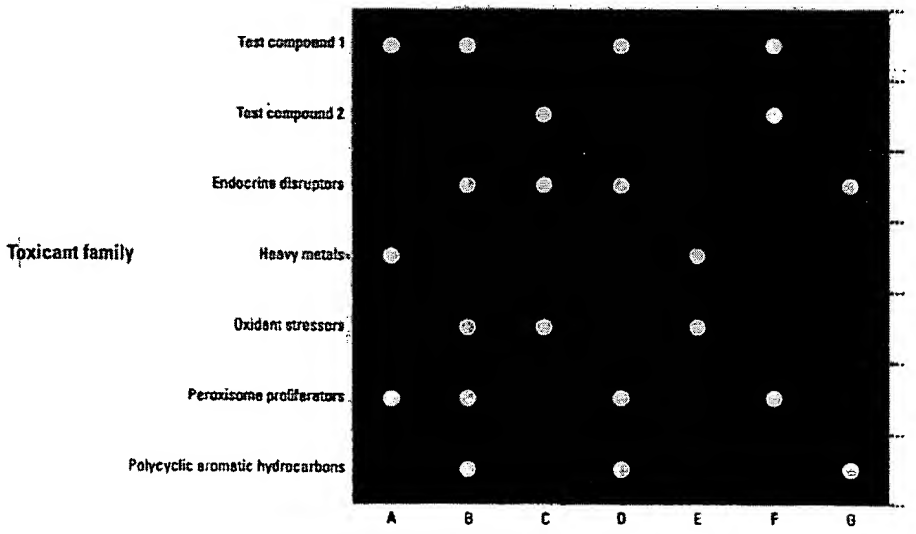


Figure 3. Gene expression profiles—also called fingerprints or signatures—of known toxicants or toxicant families may, in the future, be used to identify the potential toxicity of new drugs, etc. In this example, the genetic signature of test compound 1 is identical to that of known peroxisome proliferators, whereas that of test compound 2 does not match any known toxicant family. Based on these results, test compound 2 would be retained for further testing and test compound 1 would be eliminated.

producing chips in-house. In view of this, researchers at the RTD/NHEERL initiated the EPAMAC. This consortium brings together scientists from the EPA and a number of extramural labs with the aim of developing microarray capability through the sharing of resources and data. EPAMAC researchers are primarily interested in the developmental and toxicologic changes seen in testicular and breast tissue, and a portion of the workshop was set aside for EPAMAC members to share their ideas on how the experimental application of microarrays could facilitate their research. One of the central areas of interest to EPAMAC members is the effect of xenobiotics on male fertility and reproductive health. Of greatest concern is the effect of exposure during critical periods of development and germ cell differentiation (*9*), and how this may compromise sperm counts and quality following sexual maturation (*10*). As well as spermatogenic tissue, there is also interest in how residual mRNA found in mature sperm (*11*) could be used as an indicator of previous xenobiotic effects (it is easier to obtain a semen sample than a testicular biopsy). Arrays will be used to examine and compare the effect of exposure to heat and chemicals in testicular and epididymal gene expression profiles, with the aim of establishing relationships/associations between changes in developmental landmarks and the effects on sperm count and quality. Cluster, pattern, and other analysis of such data should help identify hidden relationships between genes that may reveal potential mechanisms of action and uncover roles for genes with unknown functions.

Summary

The full impact of DNA arrays may not be seen for several years, but the interest shown at this regional workshop indicates the high level of interest that they foster. Apart from educating and advertising the various technologies in this field, this workshop brought together a number of researchers from the Research Triangle Park area who are already using DNA arrays. The interest in sharing ideas and experiences led to the initiation of a Triangle array user's group.

Array technology is still in its infancy. This means that the hardware is still improving and there is no current consensus for standard procedures, quantitation, and interpretation. Consistency in spotting and scanning arrays is not yet optimized, and this is one of the most critical requirements of any experiment. In addition, one of the dark regions of array technology—strife in the courts over who owns what portions of it—has further muddled the future and is a potential barrier toward the development of consensus procedures.

Perhaps the greatest hurdle for the application of arrays is the actual interpretation of data. No specialists in bioinformatics attended the workshop, largely because they are rare and because as yet no one seems clear on the best method of approaching data analysis and interpretation. Cross-referencing results from multiple experiments (time, dose, repeats, different animals, different species) to identify commonly expressed genes is a great challenge. In most cases, we are still a long way from understanding how the expression of gene *X* is related to the expression of gene *Y*, and ordering gene expression to delineate causal relationships.

To the ordinary scientist in the typical laboratory, however, the most immediate problem is a lack of affordable instrumentation. One can purchase premade membranes at relatively affordable prices. Although these may be useful in identifying individual genes to pursue in more detail using other methods, the numbers that would be required for even a small routine toxicology experiment prohibit this as a truly viable approach. For the toxicologist, there is a need to carry out multiple experiments—dose responses, time curves, multiple animals, and repeats. Glass-based DNA arrays are most attractive in this context because they can be prepared in large batches from the same DNA source and accommodate control and treated samples on the same chip. Another problem with current off-the-shelf arrays is that they often do not contain one or more of the particular genes a group is interested in. One alternative is to obtain and/or produce a set of custom clones and have contract printing of membranes or slides carried out by a company such as Genomic Solutions, Inc. (Ann Arbor, MI). This approach is less expensive than laying out capital for one's own entire system, although at some point it might make economic sense to print one's own arrays.

Finally, DNA arrays are currently a team effort. They are a technology that uses a wide range of skills including engineering, statistics, molecular biology, chemistry, and bioinformatics. Because most individuals are skilled in only one or perhaps two of these areas, it appears that success with arrays may be best expected by teams of collaborators consisting of individuals having each of these skills.

Those considering array applications may be amused or goaded on by the following quote from *Fortune* magazine (*12*):

> Microprocessors have reshaped our economy, spawned vast fortunes and changed the way we live. Gene chips could be even bigger.

Although this comment may have been designed to excite the imagination rather than accurately reflect the truth, it is fair to say that the age of functional genomics is upon us. DNA arrays look set to be an important tool in this new age of biotechnology and will likely contribute answers to some of toxicology's most fundamental questions.

Speakers

Cindy Afshari, NIEHS
Linda Birnbaum, U.S. EPA
Ron Butow, University of Texas Southwestern Medical Center
Alex Chenchik, Clontech Laboratories, Inc.
David Dix, U.S. EPA
Abdel Elkahloun, Research Genetics, Inc.
Sue Fenton, U.S. EPA
Norman Hecht, University of Pennsylvania
Pat Hurban, Paradigm Genetics, Inc.
Bob Kavlock, U.S. EPA
Ernie Kawasaki, General Scanning, Inc.
Steve Krawetz, Wayne State University
Nick Mace, Genetic Microsystems, Inc.
Scott Mordecai, Affymetrix, Inc.
Kevin Morgan, Glaxo Wellcome, Inc.
Elaine Poplin, Research Genetics, Inc.
Don Rose, Cartesian Technologies, Inc.
Jim Samet, U.S. EPA
Sam Ward, University of Arizona
Jeff Welch, U.S. EPA
Reen Wu, University of California at Davis
Tim Zacharewski, Michigan State University

References and Notes

1. The chipping forecast. Nat Genet 21(Suppl 1):3–60 (1999).
2. National Center for Biotechnology Information. The Unigene System. Available: www.ncbi.nlm.nih.gov/Schuler/UniGene [cited 22 March 1999].
3. Brown PO. The Brown Lab. Available: http://cmgm.Stanford.edu/pbrown [cited 22 March 1999].
4. Chen JJ, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF, Chang F, et al. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics 51:313–324 (1998).
5. Ward S. DNA Microarray Technology to Identify Genes Controlling Spermatogenesis. Available: www.mcb.arizona.edu/wardlab/microarray.html [cited 22 March 1999].
6. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, et al. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med 4:1293–1301 (1998).
7. Brown PO. The Full Yeast Genome on a Chip. Available: http://cmgm.stanford.edu/pbrown/yeastchip.html [cited 22 March 1999].
8. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 24(3):153–159 (1999).
9. Hecht NB. Molecular mechanisms of male germ cell differentiation. Bioessays 20:555–561 (1998).
10. Zacharewski TR. Timothy R. Zacharewski. Available: www.bch.msu.edu/faculty/zachar.htm [cited 22 March 1999].
11. Kramer JA, Krawetz SA. RNA in spermatozoa: implications for the alternative haploid genome. Mol Hum Reprod 3:473–478 (1997).
12. Stipp D. Gene chip breakthrough. Fortune, March 31:56–73 (1997).
13. Kawasaki E (General Scanning Instruments, Inc., Watertown, MA). Unpublished data.
14. Flowers P, Overbeck J, Mace ML Jr, Pagliughi FM, Eggers WJE, Yonkers H, Honkanen P, Montagu J, Rose SD. Development and Performance of a Novel Microarraying System Based on Surface Tension Forces. Available: http://www.geneticmicro.com/resources/html/coldspring.html [cited 22 March 1999].
15. Butow R (University of Texas Medical Center, Dallas, TX). Unpublished data.

Subject: RE: [Fwd: Toxicology Chip]
Date: Mon. 3 Jul 2000 08:09:45 -0400
From: "Afshari.Cynthia" <afshari@niehs.nih.gov>
To: "'Diana Hamlet-Cox'" <dianahc@incyte.com>

You can see the list of clones that we have on our 12K chip at
http://manuel.niehs.nih.gov/maps/guest/clonesrch.cfm
We selected a subset of genes (2000K) that we believed critical to tox response and basic cellular processes and added a set of clones and ESTs to this. We have included a set of control genes (80-) that were selected by the NHGRI because they did not change across a large set of array experiments. However, we have found that some of these genes change signficantly after tox treatments and are in the process of looking at the variation of each of these 80+ genes across our experiments.
Our chips are constantly changing and being updated and we hope that our data will lead us to what the toxchip should really be.
I hope this answers your question.
Cindy Afshari

> ----------
> *From: Diana Hamlet-Cox*
> *Sent: Monday, June 26, 2000 8:52 PM*
> *To: afshari@niehs.nih.gov*
> *Subject: [Fwd: Toxicology Chip]*
>
> *Dear Dr. Afshari,*
>
> *Since I have not yet had a response from Bill Grigg, perhaps he was not the right person to contact.*
>
> *Can you help me in this matter? I don't need to know the sequences, necessarily, but I would like very much to know what types of sequences are being used, e.g., GPCRs (more specific?), ion channels, etc.*
>
> *Diana Hamlet-Cox*
>
> *-------- Original Message --------*
> *Subject: Toxicology Chip*
> *Date: Mon, 19 Jun 2000 18:31:48 -0700*
> *From: Diana Hamlet-Cox <dianahc@incyte.com>*
> *Organization: Incyte Pharmaceuticals*
> *To: grigg@niehs.nih.gov*
>
> *Dear Colleague:*
>
> *I am doing literature research on the use of expressed genes as pharmacotoxicology markers, and found the Press Release dated February 29, 2000 regarding the work of the NIEHS in this area. I would like to know if there is a resource I can access (or you could provide?) that would give me a list of the 12,000 genes that are on your Human ToxChip Microarray. In particular, I am interested in the criteria used to select sequences for the ToxChip, including any control sequences included in the microarray.*
>
> *Thank you for your assistance in this request.*
>
> *Diana Hamlet-Cox, Ph.D.*
> *Incyte Genomics, Inc.*
>
> --
>
> ==========

> this email message is for the sole use of the intended recipient s and
> may contain confidential and privileged information subject to
> attorney-client privilege. Any unauthorized review, use, disclosure or
> distribution is prohibited. If you are not the intended recipient,
> please contact the sender by reply email and destroy all copies of the
> original message.
> ==========
>
>

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6073–6078, May 1998
Biochemistry

Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships

STEVEN E. BRENNER*†‡, CYRUS CHOTHIA*, AND TIM J. P. HUBBARD§

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and §Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, United Kingdom

Communicated by David R. Davies, National Institute of Diabetes, Bethesda, MD, March 16, 1998 (received for review November 12, 1997)

ABSTRACT **Pairwise sequence comparison methods have been assessed using proteins whose relationships are known reliably from their structures and functions, as described in the SCOP database [Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia C. (1995) *J. Mol. Biol.* 247, 536–540]. The evaluation tested the programs BLAST [Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* 215, 403–410], WU-BLAST2 [Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460–480], FASTA [Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448], and SSEARCH [Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195–197] and their scoring schemes. The error rate of all algorithms is greatly reduced by using statistical scores to evaluate matches rather than percentage identity or raw scores. The E-value statistical scores of SSEARCH and FASTA are reliable: the number of false positives found in our tests agrees well with the scores reported. However, the P-values reported by BLAST and WU-BLAST2 exaggerate significance by orders of magnitude. SSEARCH, FASTA ktup = 1, and WU-BLAST2 perform best, and they are capable of detecting almost all relationships between proteins whose sequence identities are >30%. For more distantly related proteins, they do much less well; only one-half of the relationships between proteins with 20–30% identity are found. Because many homologs have low sequence similarity, most distant relationships cannot be detected by any pairwise comparison method; however, those which are identified may be used with confidence.**

Sequence database searching plays a role in virtually every branch of molecular biology and is crucial for interpreting the sequences issuing forth from genome projects. Given the method's central role, it is surprising that overall and relative capabilities of different procedures are largely unknown. It is difficult to verify algorithms on sample data because this requires large data sets of proteins whose evolutionary relationships are known unambiguously and independently of the methods being evaluated. However, nearly all known homologs have been identified by sequence analysis (the method to be tested). Also, it is generally very difficult to know, in the absence of structural data, whether two proteins that lack clear sequence similarity are unrelated. This has meant that although previous evaluations have helped improve sequence comparison, they have suffered from insufficient, imperfectly characterized, or artificial test data. Assessment also has been problematic because high quality database sequence searching attempts to have both sensitivity (detection of homologs) and specificity (rejection of unrelated proteins); however, these complementary goals are linked such that increasing one causes the other to be reduced.

Sequence comparison methodologies have evolved rapidly, so no previously published tests has evaluated modern versions of programs commonly used. For example, parameters in BLAST (1) have changed, and WU-BLAST2 (2)—which produces gapped alignments—has become available. The latest version of FASTA (3) previously tested was 1.6, but the current release (version 3.0) provides fundamentally different results in the form of statistical scoring.

The previous reports also have left gaps in our knowledge. For example, there has been no published assessment of thresholds for scoring schemes more sophisticated than percentage identity. Thus, the widely discussed statistical scoring measures have never actually been evaluated on large databases of real proteins. Moreover, the different scoring schemes commonly in use have not been compared.

Beyond these issues, there is a more fundamental question: in an absolute sense, how well does pairwise sequence comparison work? That is, what fraction of homologous proteins can be detected using modern database searching methods?

In this work, we attempt to answer these questions and to overcome both of the fundamental difficulties that have hindered assessment of sequence comparison methodologies. First, we use the set of distant evolutionary relationships in the SCOP: Structural Classification of Proteins database (4), which is derived from structural and functional characteristics (5). The SCOP database provides a uniquely reliable set of homologs, which are known independently of sequence comparison. Second, we use an assessment method that jointly measures both sensitivity and specificity. This method allows straightforward comparison of different sequence searching procedures. Further, it can be used to aid interpretation of real database searches and thus provide optimal and reliable results.

Previous Assessments of Sequence Comparison. Several previous studies have examined the relative performance of different sequence comparison methods. The most encompassing analyses have been by Pearson (6, 7), who compared the three most commonly used programs. Of these, the Smith–Waterman algorithm (8) implemented in SSEARCH (3) is the oldest and slowest but the most rigorous. Modern heuristics have provided BLAST (1) the speed and convenience to make it the most popular program. Intermediate between these two is FASTA (3), which may be run in two modes offering either greater speed (ktup = 2) or greater effectiveness (ktup = 1). Pearson also considered different parameters for each of these programs.

To test the methods, Pearson selected two representative proteins from each of 67 protein superfamilies defined by the PIR database (9). Each was used as a query to search the database, and the matched proteins were marked as being homologous or unrelated according to their membership of PIR

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EPQ, errors per query.

†Present address: Department of Structural Biology, Stanford University, Fairchild Building D-109, Stanford, CA 94305-5126
‡To whom reprints requests should be addressed. e-mail: brenner@hyper.stanford.edu.

superfamilies. Pearson found that modern matrices and "ln-scaling" of raw scores improve results considerably. He also reported that the rigorous Smith–Waterman algorithm worked slightly better than FASTA, which was in turn more effective than BLAST.

Very large scale analyses of matrices have been performed (10), and Henikoff and Henikoff (11) also evaluated the effectiveness of BLAST and FASTA. Their test with BLAST considered the ability to detect homologs above a predetermined score but had no penalty for methods which also reported large numbers of spurious matches. The Henikoffs searched the SWISS-PROT database (12) and used PROSITE (13) to define homologous families. Their results showed that the BLOSUM62 matrix (14) performed markedly better than the extrapolated PAM-series matrices (15), which previously had been popular.

A crucial aspect of any assessment is the data that are used to test the ability of the program to find homologs. But in Pearson's and the Henikoffs' evaluations of sequence comparison, the correct results were effectively unknown. This is because the superfamilies in PIR and PROSITE are principally created by using the same sequence comparison methods which are being evaluated. Interdependency of data and methods creates a "chicken and egg" problem, and means for example, that new methods would be penalized for correctly identifying homologs missed by older programs. For instance, immunoglobulin variable and constant domains are clearly homologous, but PIR places them in different superfamilies. The problem is widespread: each superfamily in PIR 48.00 with a structural homolog is itself homologous to an average of 1.6 other PIR superfamilies (16).

To surmount these sorts of difficulties, Sander and Schneider (17) used protein structures to evaluate sequence comparison. Rather than comparing different sequence comparison algorithms, their work focused on determining a length-dependent threshold of percentage identity, above which all proteins would be of similar structure. A result of this analysis was the HSSP equation; it states that proteins with 25% identity over 80 residues will have similar structures, whereas shorter alignments require higher identity. (Other studies also have used structures (18–20), but these focused on a small number of model proteins and were principally oriented toward evaluating alignment accuracy rather than homology detection.)

A general solution to the problem of scoring comes from statistical measures (i.e., E-values and P-values) based on the extreme value distribution (21). Extreme value scoring was implemented analytically in the BLAST program using the Karlin and Altschul statistics (22, 23) and empirical approaches have been recently added to FASTA and SSEARCH. In addition to being heralded as a reliable means of recognizing significantly similar proteins (24, 25), the mathematical tractability of statistical scores "is a crucial feature of the BLAST algorithm" (1). The validity of this scoring procedure has been tested analytically and empirically (see ref. 2 and references in ref. 24). However, all large empirical tests used random sequences that may lack the subtle structure found within biological sequences (26, 27) and obviously do not contain any real homologs. Thus, although many researchers have suggested that statistical scores be used to rank matches (24, 25, 28), there have been no large rigorous experiments on biological data to determine the degree to which such rankings are superior.

A Database for Testing Homology Detection. Since the discovery that the structures of hemoglobin and myoglobin are very similar though their sequences are not (29), it has been apparent that comparing structures is a more powerful (if less convenient) way to recognize distant evolutionary relationships than comparing sequences. If two proteins show a high degree of similarity in their structural details and function, it is very probable that they have an evolutionary relationship though their sequence similarity may be low.

The recent growth of protein structure information combined with the comprehensive evolutionary classification in the SCOP database (4, 5) have allowed us to overcome previous limitations. With these data, we can evaluate the performance of sequence comparison methods on real protein sequences whose relationships are known confidently. The SCOP database uses structural information to recognize distant homologs, the large majority of which can be determined unambiguously. These superfamilies, such as the globins or the immunoglobulins, would be recognized as related by the vast majority of the biological community despite the lack of high sequence similarity.

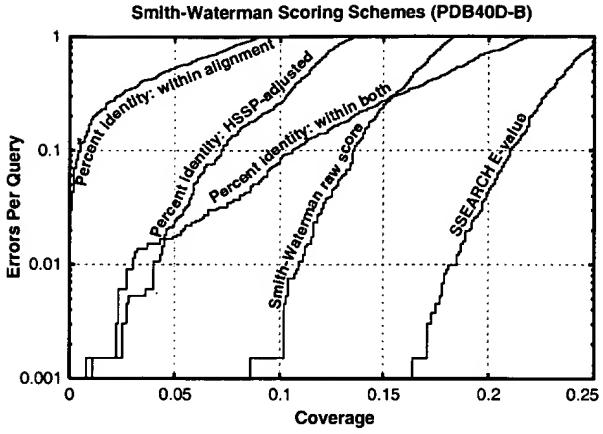
From SCOP, we extracted the sequences of domains of proteins in the Protein Data Bank (PDB) (30) and created two databases. One (PDB90D-B) has domains, which were all <90% identical to any other, whereas (PDB40D-B) had those <40% identical. The databases were created by first sorting all protein domains in SCOP by their quality and making a list. The highest quality domain was selected for inclusion in the database and removed from the list. Also removed from the list (and discarded) were all other domains above the threshold level of identity to the selected domain. This process was repeated until the list was empty. The PDB40D-B database contains 1,323 domains, which have 9,044 ordered pairs of distant relationships, or ≈0.5% of the total 1,749,006 ordered pairs. In PDB90D-B, the 2,079 domains have 53,988 relationships, representing 1.2% of all pairs. Low complexity regions of sequence can achieve spurious high scores, so these were masked in both databases by processing with the SEG program (27) using recommended parameters: 12 1.8 2.0. The databases used in this paper are available from http://sss.stanford.edu/sss/, and databases derived from the current version of SCOP may be found at http://scop.mrc-lmb.cam.ac.uk/scop/.

Analyses from both databases were generally consistent, but PDB40D-B focuses on distantly related proteins and reduces the heavy overrepresentation in the PDB of a small number of families (31, 32), whereas PDB90D-B (with more sequences) improves evaluations of statistics. Except where noted otherwise, the distant homolog results here are from PDB40D-B. Although the precise numbers reported here are specific to the structural domain databases used, we expect the trends to be general.

Assessment Data and Procedure. Our assessment of sequence comparison may be divided into four different major categories of tests. First, using just a single sequence comparison algorithm at a time, we evaluated the effectiveness of different scoring schemes. Second, we assessed the reliability of scoring procedures, including an evaluation of the validity of statistical scoring. Third, we compared sequence comparison algorithms (using the optimal scoring scheme) to determine their relative performance. Fourth, we examined the distribution of homologs and considered the power of pairwise sequence comparison to recognize them. All of the analyses used the databases of structurally identified homologs and a new assessment criterion.

The analyses tested BLAST (1), version 1.4.9MP, and WU-BLAST2 (2), version 2.0a13MP. Also assessed was the FASTA package, version 3.0t76 (3), which provided FASTA and the SSEARCH implementation of Smith–Waterman (8). For SSEARCH and FASTA, we used BLOSUM45 with gap penalties −12/−1 (7, 16). The default parameters and matrix (BLOSUM62) were used for BLAST and WU-BLAST2.

The "Coverage Vs. Error" Plot. To test a particular protocol (comprising a program and scoring scheme), each sequence from the database was used as a query to search the database. This yielded ordered pairs of query and target sequences with associated scores, which were sorted, on the basis of their scores, from best to worst. The ideal method would have



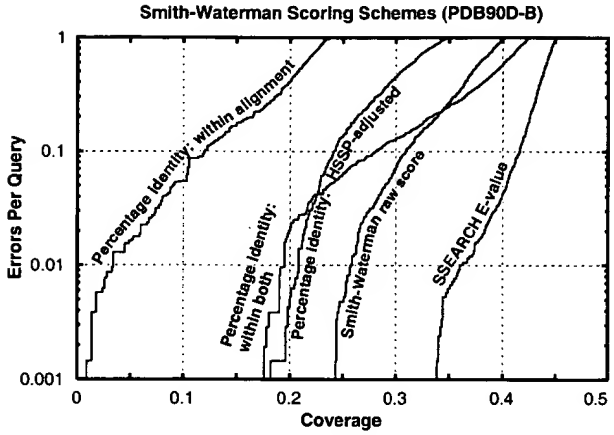


FIG. 1. Coverage vs. error plots of different scoring schemes for SSEARCH Smith–Waterman. (*A*) Analysis of PDB40D-B database. (*B*) Analysis of PDB90D-B database. All of the proteins in the database were compared with each other using the SSEARCH program. The results of this single set of comparisons were considered using five different scoring schemes and assessed. The graphs show the coverage and errors per query (EPQ) for statistical scores, raw scores, and three measures using percentage identity. In the coverage vs. error plot, the x axis indicates the fraction of all homologs in the database (known from structure) which have been detected. Precisely, it is the number of detected pairs of proteins with the same fold divided by the total number of pairs from a common superfamily. PDB40D-B contains a total of 9,044 homologs, so a score of 10% indicates identification of 904 relationships. The y axis reports the number of EPQ. Because there are 1,323 queries made in the PDB40D-B all-vs.-all comparison, 13 errors corresponds to 0.01, or 1% EPQ. The y axis is presented on a log scale to show results over the widely varying degrees of accuracy which may be desired. The scores that correspond to the levels of EPQ and coverage are shown in Fig. 4 and Table 1. The graph demonstrates the trade-off between sensitivity and selectivity. As more homologs are found (moving to the right), more errors are made (moving up). The ideal method would be in the lower right corner of the graph, which corresponds to identifying many evolutionary relationships without selecting unrelated proteins. Three measures of percentage identity are plotted. Percentage identity within alignment is the degree of identity within the aligned region of the proteins, without consideration of the alignment length. Percentage identity within both is the number of identical residues in the aligned region as a percentage of the average length of the query and target proteins. The HSSP equation (17) is $H = 290.15l^{-0.562}$ where l is length for $10 < l < 80$; $H > 100$ for $l < 10$; $H = 24.7$ for $l > 80$. The percentage identity HSSP-adjusted score is the percent identity within the alignment minus H. Smith–Waterman raw scores and E-values were taken directly from the sequence comparison program.

perfect separation, with all of the homologs at the top of the list and unrelated proteins below. In practice, perfect separation is impossible to achieve so instead one is interested in drawing a threshold above which there are the largest number of related pairs of sequences consistent with an acceptable error rate.

Our procedure involved measuring the coverage and error for every threshold. Coverage was defined as the fraction of structurally determined homologs that have scores above the selected threshold; this reflects the sensitivity of a method. Errors per query (EPQ), an indicator of selectivity, is the number of nonhomologous pairs above the threshold divided by the number of queries. Graphs of these data, called coverage vs. error plots, were devised to understand how protocols compare at different levels of accuracy. These graphs share effectively all of the beneficial features of Reciever Operating Characteristic (ROC) plots (33, 34) but better represent the high degrees of accuracy required in sequence comparison and the huge background of nonhomologs.

This assessment procedure is directly relevant to practical sequence database searching, for it provides precisely the information necessary to perform a reliable sequence database search. The EPQ measure places a premium on score consistency; that is, it requires scores to be comparable for different queries. Consistency is an aspect which has been largely

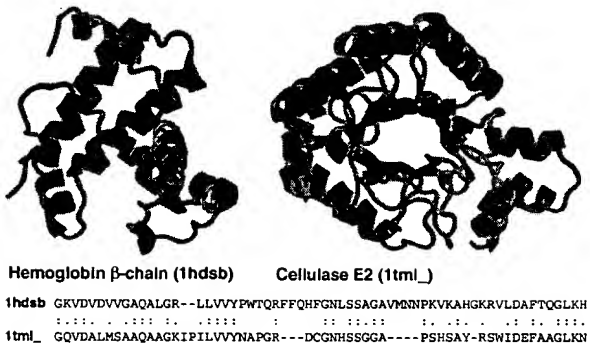


FIG. 2. Unrelated proteins with high percentage identity. Hemoglobin β-chain (PDB code 1hds chain b, ref. 38, *Left*) and cellulase E2 (PDB code 1tml, ref. 39, *Right*) have 39% identity over 64 residues, a level which is often believed to be indicative of homology. Despite this high degree of identity, their structures strongly suggest that these proteins are not related. Appropriately, neither the raw alignment score of 85 nor the E-value of 1.3 is significant. Proteins rendered by RASMOL (40).

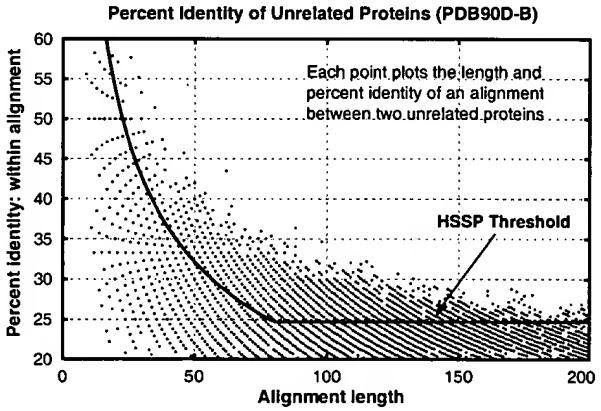


FIG. 3. Length and percentage identity of alignments of unrelated proteins in PDB90D-B: Each pair of nonhomologous proteins found with SSEARCH is plotted as a point whose position indicates the length and the percentage identity within the alignment. Because alignment length and percentage identity are quantized, many pairs of proteins may have exactly the same alignment length and percentage identity. The line shows the HSSP threshold (though it is intended to be applied with a different matrix and parameters).

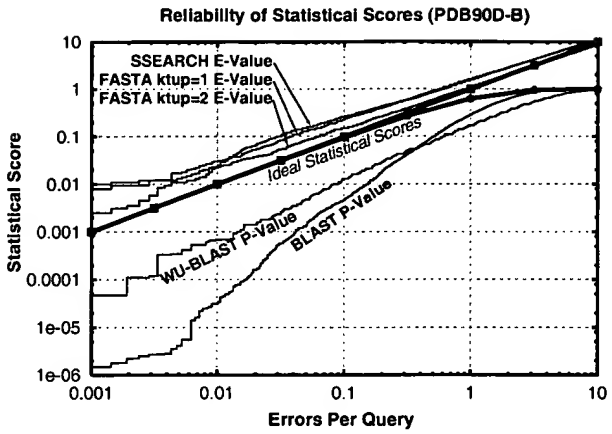


FIG. 4. Reliability of statistical scores in PDB90D-B: Each line shows the relationship between reported statistical score and actual error rate for a different program. E-values are reported for SSEARCH and FASTA, whereas P-values are shown for BLAST and WU-BLAST2. If the scoring were perfect, then the number of errors per query and the E-values would be the same, as indicated by the upper bold line. (P-values should be the same as EPQ for small numbers, and diverges at higher values, as indicated by the lower bold line.) E-values from SSEARCH and FASTA are shown to have good agreement with EPQ but underestimate the significance slightly. BLAST and WU-BLAST2 are overconfident, with the degree of exaggeration dependent upon the score. The results for PDB40D-B were similar to those for PDB90D-B despite the difference in number of homologs detected. This graph could be used to roughly calibrate the reliability of a given statistical score.

ignored in previous tests but is essential for the straightforward or automatic interpretation of sequence comparison results. Further, it provides a clear indication of the confidence that should be ascribed to each match. Indeed, the EPQ measure should approximate the expectation value reported by database searching programs, if the programs' estimates are accurate.

The Performance of Scoring Schemes. All of the programs tested could provide three fundamental types of scores. The first score is the percentage identity, which may be computed in several ways based on either the length of the alignment or the lengths of the sequences. The second is a "raw" or "Smith–Waterman" score, which is the measure optimized by the Smith–Waterman algorithm and is computed by summing the substitution matrix scores for each position in the alignment and subtracting gap penalties. In BLAST, a measure related to this score is scaled into bits. Third is a statistical score based on the extreme value distribution. These results are summarized in Fig. 1.

Sequence Identity. Though it has been long established that percentage identity is a poor measure (35), there is a common rule-of-thumb stating that 30% identity signifies homology. Moreover, publications have indicated that 25% identity can be used as a threshold (17, 36). We find that these thresholds, originally derived years ago, are not supported by present results. As databases have grown, so have the possibilities for chance alignments with high identity; thus, the reported cutoffs lead to frequent errors. Fig. 2 shows one of the many pairs of proteins with very different structures that nonetheless have high levels of identity over considerable aligned regions. Despite the high identity, the raw and the statistical scores for such incorrect matches are typically not significant. The principal reasons percentage identity does so poorly seem to be that it ignores information about gaps and about the conservative or radical nature of residue substitutions.

From the PDB90D-B analysis in Fig. 3, we learn that 30% identity is a reliable threshold for this database only for sequence alignments of at least 150 residues. Because one unrelated pair of proteins has 43.5% identity over 62 residues, it is probably necessary for alignments to be at least 70 residues in length before 40% is a reasonable threshold, for a database of this particular size and composition.

At a given reliability, scores based on percentage identity detect just a fraction of the distant homologs found by statistical scoring. If one measures the percentage identity in the aligned regions without consideration of alignment length, then a negligible number of distant homologs are detected. Use of the HSSP equation improves the value of percentage identity, but even this measure can find only 4% of all known homologs at 1% EPQ. In short, percentage identity discards most of the information measured in a sequence comparison.

Raw Scores. Smith–Waterman raw scores perform better than percentage identity (Fig. 1), but ln-scaling (7) provided no notable benefit in our analysis. It is necessary to be very precise when using either raw or bit scores because a 20% change in cutoff score could yield a tenfold difference in EPQ. However, it is difficult to choose appropriate thresholds because the reliability of a bit score depends on the lengths of the proteins matched and the size of the database. Raw score thresholds also are affected by matrix and gap parameters.

Statistical Scores. Statistical scores were introduced partly to overcome the problems that arise from raw scores. This scoring scheme provides the best discrimination between homologous proteins and those which are unrelated. Most

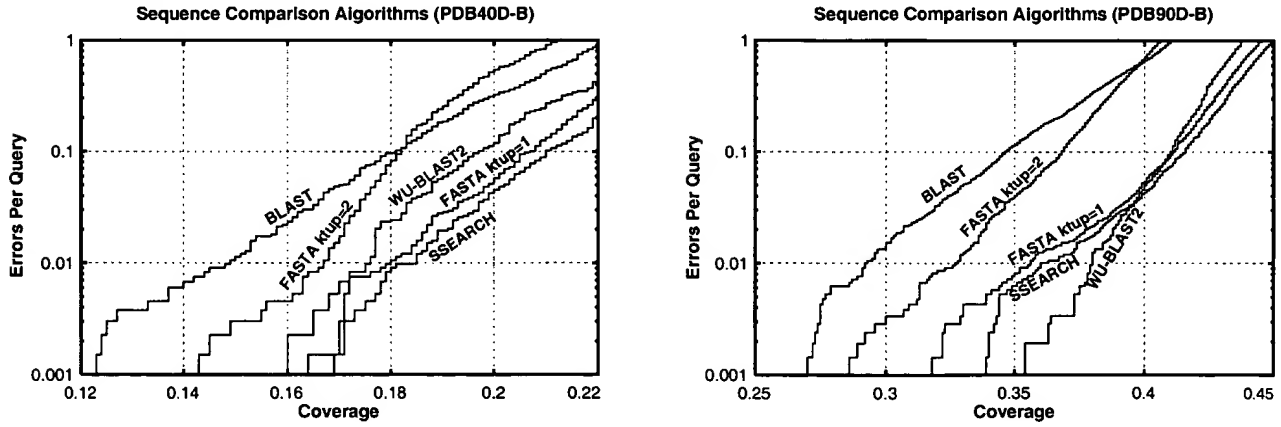


FIG. 5. Coverage vs. error plots of different sequence comparison methods: Five different sequence comparison methods are evaluated, each using statistical scores (E- or P-values). (*A*) PDB40D-B database. In this analysis, the best method is the slow SSEARCH, which finds 18% of relationships at 1% EPQ. FASTA ktup = 1 and WU-BLAST2 are almost as good. (*B*) PDB90D-B database. The quick WU-BLAST2 program provides the best coverage at 1% EPQ on this database, although at higher levels of error it becomes slightly worse than FASTA ktup = 1 and SSEARCH.

likely, its power can be attributed to its incorporation of more information than any other measure; it takes account of the full substitution and gap data (like raw scores) but also has details about the sequence lengths and composition and is scaled appropriately.

We find that statistical scores are not only powerful, but also easy to interpret. SSEARCH and FASTA show close agreement between statistical scores and actual number of errors per query (Fig. 4). The expectation value score gives a good, slightly conservative estimate of the chances of the two sequences being found at random in a given query. Thus, an E-value of 0.01 indicates that roughly one pair of nonhomologs of this similarity should be found in every 100 different queries. Neither raw scores nor percentage identity can be interpreted in this way, and these results validate the suitability of the extreme value distribution for describing the scores from a database search.

The P-values from BLAST also should be directly interpretable but were found to overstate significance by more than two orders of magnitude for 1% EPQ for this database. Nonetheless, these results strongly suggest that the analytic theory is fundamentally appropriate. WU-BLAST2 scores were more reliable than those from BLAST, but also exaggerate expected confidence by more than an order of magnitude at 1% EPQ.

Overall Detection of Homologs and Comparison of Algorithms. The results in Fig. 5*A* and Table 1 show that pairwise sequence comparison is capable of identifying only a small fraction of the homologous pairs of sequences in PDB40D-B. Even SSEARCH with E-values, the best protocol tested, could find only 18% of all relationships at a 1% EPQ. BLAST, which identifies 15%, was the worst performer, whereas FASTA ktup = 1 is nearly as effective as SSEARCH. FASTA ktup = 2 and WU-BLAST2 are intermediate in their ability to detect homologs. Comparison of different algorithms indicates that those capable of identifying more homologs are generally slower. SSEARCH is 25 times slower than BLAST and 6.5 times slower than FASTA ktup = 1. WU-BLAST2 is slightly faster than FASTA ktup = 2, but the latter has more interpretable scores.

In PDB90D-B, where there are many close relationships, the best method can identify only 38% of structurally known homologs (Fig. 5*B*). The method which finds that many relationships is WU-BLAST2. Consequently, we infer that the differences between FASTA kup = 1, SSEARCH, and WU-BLAST2 programs are unlikely to be significant when compared with variation in database composition and scoring reliability.

Fig. 6 helps to explain why most distant homologs cannot be found by sequence comparison: a great many such relationships have no more sequence identity than would be expected by chance. SSEARCH with E-values can recognize >90% of the homologous pairs with 30–40% identity. In this region, there are 30 pairs of homologous proteins that do not have significant E-values, but 26 of these involve sequences with <50 residues. Of sequences having 25–30% identity, 75% are identified by SSEARCH E-values. However, although the number of homologs grows at lower levels of identity, the detection falls off sharply: only 40% of homologs with 20–25% identity

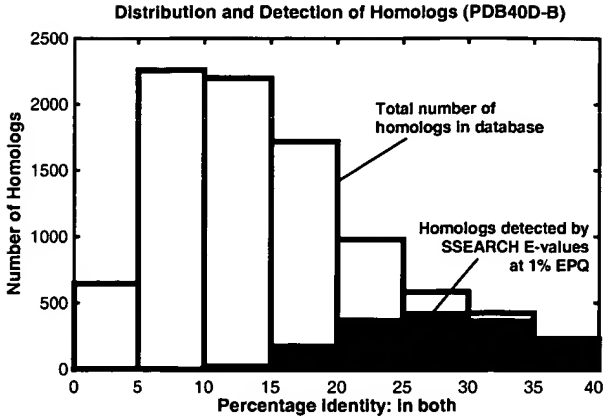


FIG. 6. Distribution and detection of homologs in PDB40D-B. Bars show the distribution of homologous pairs PDB40D-B according to their identity (using the measure of identity in both). Filled regions indicate the number of these pairs found by the best database searching method (SSEARCH with E-values) at 1% EPQ. The PDB40D-B database contains proteins with <40% identity, and as shown on this graph, most structurally identified homologs in the database have diverged extremely far in sequence and have <20% identity. Note that the alignments may be inaccurate, especially at low levels of identity. Filled regions show that SSEARCH can identify most relationships that have 25% or more identity, but its detection wanes sharply below 25%. Consequently, the great sequence divergence of most structurally identified evolutionary relationships effectively defeats the ability of pariwise sequence comparison to detect them.

are detected and only 10% of those with 15–20% can be found. These results show that statistical scores can find related proteins whose identity is remarkably low; however, the power of the method is restricted by the great divergence of many protein sequences.

After completion of this work, a new version of pairwise BLAST was released: BLASTGP (37). It supports gapped alignments, like WU-BLAST2, and dispenses with sum statistics. Our initial tests on BLASTGP using default parameters show that its E-values are reliable and that its overall detection of homologs was substantially better than that of ungapped BLAST, but not quite equal to that of WU-BLAST2.

CONCLUSION

The general consensus amongst experts (see refs. 7, 24, 25, 27 and references therein) suggests that the most effective sequence searches are made by (*i*) using a large current database in which the protein sequences have been complexity masked and (*ii*) using statistical scores to interpret the results. Our experiments fully support this view.

Our results also suggest two further points. First, the E-values reported by FASTA and SSEARCH give fairly accurate estimates of the significance of each match, but the P-values provided by BLAST and WU-BLAST2 underestimate the true

Table 1. Summary of sequence comparison methods with PDB40D-B

Method	Relative Time*	1% EPQ Cutoff	Coverage at 1% EPQ
SSEARCH % identity: within alignment	25.5	>70%	<0.1
SSEARCH % identity: within both	25.5	34%	3.0
SSEARCH % identity: HSSP-scaled	25.5	35% (HSSP + 9.8)	4.0
SSEARCH Smith–Waterman raw scores	25.5	142	10.5
SSEARCH E-values	25.5	0.03	18.4
FASTA ktup = 1 E-values	3.9	0.03	17.9
FASTA ktup = 2 E-values	1.4	0.03	16.7
WU-BLAST2 P-values	1.1	0.003	17.5
BLAST P-values	1.0	0.00016	14.8

*Times are from large database searches with genome proteins.

extent of errors. Second, SSEARCH, WU-BLAST2, and FASTA ktup = 1 perform best, though BLAST and FASTA ktup = 2 detect most of the relationships found by the best procedures and are appropriate for rapid initial searches.

The homologous proteins that are found by sequence comparison can be distinguished with high reliability from the huge number of unrelated pairs. However, even the best database searching procedures tested fail to find the large majority of distant evolutionary relationships at an acceptable error rate. Thus, if the procedures assessed here fail to find a reliable match, it does not imply that the sequence is unique; rather, it indicates that any relatives it might have are distant ones.**

**Additional and updated information about this work, including supplementary figures, may be found at http://sss.stanford.edu/sss/.

The authors are grateful to Drs. A. G. Murzin, M. Levitt, S. R. Eddy, and G. Mitchison for valuable discussion. S.E.B. was principally supported by a St. John's College (Cambridge, UK) Benefactors' Scholarship and by the American Friends of Cambridge University. S.E.B. dedicates his contribution to the memory of Rabbi Albert T. and Clara S. Bilgray.

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
2. Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* **266,** 460–480.
3. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
4. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540.
5. Brenner, S. E., Chothia, C., Hubbard, T. J. P. & Murzin, A. G. (1996) *Methods Enzymol.* **266,** 635–643.
6. Pearson, W. R. (1991) *Genomics* **11,** 635–650.
7. Pearson, W. R. (1995) *Protein Sci.* **4,** 1145–1160.
8. Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* **147,** 195–197.
9. George, D. G., Hunt, L. T. & Barker, W. C. (1996) *Methods Enzymol.* **266,** 41–59.
10. Vogt, G., Etzold, T. & Argos, P. (1995) *J. Mol. Biol.* **249,** 816–831.
11. Henikoff, S. & Henikoff, J. G. (1993) *Proteins* **17,** 49–61.
12. Bairoch, A. & Apweiler, R. (1996) *Nucleic Acids Res.* **24,** 21–25.
13. Bairoch, A., Bucher, P. & Hofmann, K. (1996) *Nucleic Acids Res.* **24,** 189–196.
14. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10915–10919.
15. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure,* ed. Dayhoff, M. (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
16. Brenner, S. E. (1996) Ph.D. thesis. (University of Cambridge, UK).
17. Sander, C. & Schneider, R. (1991) *Proteins* **9,** 56–68.
18. Johnson, M. S. & Overington, J. P. (1993) *J. Mol. Biol.* **233,** 716–738.
19. Barton, G. J. & Sternberg, M. J. E. (1987) *Protein Eng.* **1,** 89–94.
20. Lesk, A. M., Levitt, M. & Chothia, C. (1986) *Protein Eng.* **1,** 77–78.
21. Arratia, R., Gordon, L. & M, W. (1986) *Ann. Stat.* **14,** 971–993.
22. Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2264–2268.
23. Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5873–5877.
24. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6,** 119–129.
25. Pearson, W. R. (1996) *Methods Enzymol.* **266,** 227–258.
26. Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman, M. S. (1984) *Nucleic Acids Res.* **12,** 215–226.
27. Wootton, J. C. & Federhen, S. (1996) *Methods Enzymol.* **266,** 554–571.
28. Waterman, M. S. & Vingron, M. (1994) *Stat. Science* **9,** 367–381.
29. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13,** 669–678.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications,* eds. Allen, F. H., Bergerhoff, G. & Sievers, R. (Data Comm. Intl. Union Crystallogr., Cambridge, UK), pp. 107–132.
31. Brenner, S. E., Chothia, C. & Hubbard, T. J. P. (1997) *Curr. Opin. Struct. Biol.* **7,** 369–376.
32. Orengo, C., Michie, A., Jones S, Jones D. T, Swindells M. B. & Thornton, J. (1997) *Structure (London)* **5,** 1093–1108.
33. Zweig, M. H. & Campbell, G. (1993) *Clin. Chem.* **39,** 561–577.
34. Gribskov, M. & Robinson, N. L. (1996) *Comput. Chem.* **20,** 25–33.
35. Fitch, W. M. (1966) *J. Mol. Biol.* **16,** 9–16.
36. Chung, S. Y. & Subbiah, S. (1996) *Structure (London)* **4,** 1123–1127.
37. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
38. Girling, R., Schmidt, W., Jr, Houston, T., Amma, E. & Huisman, T. (1979) *J. Mol. Biol.* **131,** 417–433.
39. Spezio, M., Wilson, D. & Karplus, P. (1993) *Biochemistry* **32,** 9906–9916
40. Sayle, R. A. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20,** 374–376.

Docket No.: PF-0636 RCE
USSN: 09/831,458
Ref. No. 8 of 17

HIV gp120 receptors on human dendritic cells

Stuart G. Turville, Jim Arthos, Kelli Mac Donald, Garry Lynch, Hassan Naif, Georgina Clark, Derek Hart, and Anthony L. Cunningham

Dendritic cells (DCs) are important targets for human immunodeficiency virus (HIV) because of their roles during transmission and also maintenance of immune competence. Furthermore, DCs are a key cell in the development of HIV vaccines. In both these settings the mechanism of binding of the HIV envelope protein gp120 to DCs is of importance. Recently a single C-type lectin receptor (CLR), DC-SIGN, has been reported to be the predominant receptor on monocyte-derived DCs (MDDCs) rather than CD4. In this study a novel biotinylated gp120 assay was used to determine whether CLR or CD4 were predominant receptors on MDDCs and ex vivo blood DCs. CLR bound more than 80% of gp120 on MDDCs, with residual binding attributable to CD4, reconfirming that CLRs were the major receptors for gp120 on MDDCs. However, in contrast to recent reports, gp120 binding to at least 3 CLRs was observed: DC-SIGN, mannose receptor, and unidentified trypsin resistant CLR(s). In marked contrast, freshly isolated and cultured CD11c^{+ve} and CD11c^{-ve} blood DCs only bound gp120 via CD4. In view of these marked differences between MDDCs and blood DCs, HIV capture by DCs and transfer mechanisms to T cells as well as potential antigenic processing pathways will need to be determined for each DC phenotype. (Blood. 2001;98:2482-2488)

Introduction

Dendritic cells (DCs) play a major role in human immunodeficiency virus (HIV) pathogenesis. Peripheral or surveillance mucosal DCs are one of the first cell types infected and are distributed in the vaginal, ectocervical, and anal mucosa,[1,2] allowing contact with HIV during mucosal exposure. Thus, after vaginal inoculation with simian immunodeficiency virus in macaques, DCs are the predominant cell type infected.[3] Furthermore, the ability of DCs to cluster with and stimulate T cells may also play a key role in establishing infection. DCs from skin, mucosa, and blood of humans and macaques can participate in highly productive HIV and simian immunodeficiency virus infection in DC–T-cell cocultures and illustrates the importance of this natural DC–T-cell synergy.[4-7]

Key aspects of HIV binding to DC via gp120 are ill-defined, particularly to the different types of DCs. CD11c^{+ve} and CD11c^{-ve} blood DCs, Langerhans cells (LCs), and in vitro–derived monocyte-derived DCs (MDDCs) all express CD4 and CCR5 and can be productively infected in vitro.[8-12] However, HIV also bound several DC populations independently of CD4.[8,13,14] The heavy glycosylation of gp120 with mannose and fucose saccharides suggested HIV bound to cells also via lectin receptors. Binding of gp120 to a novel C-type lectin receptor (CLR), originally identified from a placental complementary DNA (cDNA) library[15] on the basis of HIV gp120 binding and named clone 11, on MDDCs was recently reported.[14,16] The adhesion properties of this CLR were also defined and the receptor subsequently renamed DC-SIGN (dendritic cell specific ICAM-3 grabbing nonintegrin). Although MDDCs express a diverse and abundant array of CLRs in addition to DC-SIGN,[16-24] and given substantial overlap in saccharide recognition by such CLRs, they may also serve as receptors for gp120 on MDDCs. The roles of CD4 and CLRs on most other in vivo DC types are unknown.

This study aimed to define the contributions of CD4 and CLRs in binding gp120, to address and identify the capacity of other CLRs including DC-SIGN during monocyte differentiation to mature MDDCs and, more importantly, to compare such populations with ex vivo blood DCs. Understanding the mechanisms of gp120 binding to different DC populations would help define the early events of HIV transmission via DCs in blood or mucosal tissue and improve intervention strategies. Definition of the mechanisms of HIV/gp120 binding and processing by DCs will also assist future HIV vaccine strategies and immunotherapy.

Materials and methods

MDDC generation and culture

Monocytes were isolated from 500 mL of blood (Parramatta Blood Bank, Australia) by countercurrent elutriation as previously described.[25,26] Monocytes were further depleted of contaminating cells by using a monocyte-enrichment cocktail (StemCell Technologies, Vancouver, BC, Canada). Monocyte fractions were at least 97% CD11c^{+ve}, at least 90% CD14^{+ve}, and 0.1% or less CD3^{+ve}. DCs were converted as previously described[27,28] using 500 U/mL interleukin-4 and 400 U/mL granulocyte-macrophage colony-stimulating factor (GM-CSF) (Schering-Plough, Kenilworth, NJ). At day 6 cells were at least 95% CD1a^{+ve}, CD11c^{+ve} with no detectable CD14, CD3, or CD83 populations. MDDCs were matured by culture for 48 hours with 10 ng/mL tumor necrosis factor α (TNF-α) (R&D Systems, Minneapolis, MN).

Isolation and culture of blood DCs

Blood DCs were isolated from 500 mL of blood (Mater Hospital, Brisbane, Australia) using Ficoll-Paque (Amersham Pharmacia-Biotech, Uppsala, Sweden). Residual erythrocytes were removed by Vitalyze as per the

From the Center For Virus Research, Westmead Millennium Institute, Sydney, Australia; National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; and Mater Medical Research Institute, Brisbane, Australia.

Submitted April 3, 2001; accepted June 14, 2001.

Supported by grants from the Westmead Millennium Foundation, Australian National Center in HIV Virology Research, and the Mater Trust.

Reprints: Anthony L. Cunningham, Westmead Millennium Institute, PO Box 412, Darcy Road, Westmead, NSW 2145, Australia; e-mail: tony_cunningham@wmi.usyd.edu.au.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

manufacturer's instructions (BioErgonomics, St Paul, MN). Peripheral blood mononuclear cells (PBMCs) were labeled with a mixture of anti-CD3 (OKT3), CD14 (CMRF31), CD11b (OKM1), CD16 (HUNK-2), and CD19 (FMC63) monoclonal antibodies (mAbs). After incubation with Biomag goat antimouse immunoglobulin–coated magnetic beads (Polysciences, Warrington, PA), labeled cells were removed by first preclearing with a MPC-1 magnet (Dynal, Oslo, Norway) and then passing through a Miltenyi cell separation column using a Variomacs magnet (Miltenyi Biotech, Gladbach, Germany). Depleted PBMCs were labeled with fluorescein isothiocyanate (FITC)–goat antimouse (Becton Dickinson, San Jose, CA) and negative cells separated by sorting on a FACSVantage (Becton Dickinson). For cultured blood DCs, DCs were incubated overnight at a concentration of 1×10^6 cells per milliliter in RPMI 1640 supplemented with 10% fetal calf serum and 10 ng/mL interleukin-3 (Gibco, Grand Island, NY) and 200 U/mL GM-CSF (Novartis, Basel, Switzerland).

HIV gp120 binding and inhibition studies

Purified HIV gp120 from the BaL isolate (courtesy of Ray Sweet, SmithKline Beecham, King of Prussia, PA) was biotinylated with EZ-Link NHS-LC-Biotin as per the manufacturer (Pierce, Rockford, IL). Biotinylation of gp120 did not affect the ability of the molecule to bind to CD4 and was confirmed in an sCD4 capture enzyme-linked immunosorbent assay with detection via streptavidin horseradish peroxidase (data not shown). In addition, nonbiotinylated gp120 material from the isolates BaL and 92MW959, using detection with purified and biotinylated human polyclonal antibodies from HIV-seropositive patients (Cellular Products, Buffalo, NY), produced equivalent results to biotinylated gp120 from respective isolates. In particular, the saturating concentrations of gp120 and the relative binding of gp120 by CD4 and CLR on MDDCs were the same by both methods. However, biotinylated gp120 binding assay was routinely used because it reduced one additional antibody staining step, reduced the variability of antibody binding, and allowed for flexibility when working with blood DCs, which are labeled with multiple antibodies for detection of multiple DC subsets.

For binding and inhibition studies, cells were preincubated for 40 minutes in binding media (RPMI 1640 without sodium bicarbonate [Gibco] with 1% bovine serum albumin and 10 mM HEPES [Calbiochem, San Diego, CA] pH 7.4) as above at 4°C with stated concentrations of inhibitors, followed by incubation with b-gp120 (2-fold the predetermined concentration for cellular saturation). Levels of inhibitors, with the exception of mAbs, were initially determined using a broad range of concentrations to assess the maximal level of gp120 blocking. In the cases of mAb, concentrations were routinely 5-fold that of cellular saturation. Cells were then washed twice, and measurement of bound b-gp120 was carried out by incubation of 1×10^6 cells (2×10^5 cells/200 μL) with 5 μg/mL streptavidin Oregon Green 488 (Molecular Probes, Eugene, OR) or avidin FITC (Becton Dickinson) and detected by flow cytometry.

Flow cytometric analysis

For surface staining, cells were treated as previously described.[29] In gp120 binding studies, cells were preincubated with b-gp120 at various concentrations for 40 minutes at 4°C in binding media. Antibodies used were CD14-phycoerythrin (PE), immunoglobulin G1 (IgG1)–PE, IgG1-FITC, CD3-FITC, IgG1, goat antimouse FITC (all from Becton Dickinson), CD83, CD86, CD1a-FITC, MR (clones 19 and 3.29), and HLA-DR–PE/P5 (all from PharMingen, San Diego, CA, except anti-MR 3.29, which is from Immunotech, Marseille, France). The CD4 mAbs used were Leu3a (Becton Dickinson), OKT4 (American Type Culture Collection, Manassas, VA), and Q4120 (a generous gift from Quentin Sattentau). The mAbs to DC-SIGN (AZN-D1 and AZN-D2) and associated experiments were a part of the 7th Leukocyte Differentiation DC Antigen Workshop (kindly donated by Yvette van Kooyk). Detection of b-gp120 and biotinylated polyclonal sera to HIV (Cellular Products) was via strepdavidin Oregon Green 488 or avidin FITC.

DC-SIGN reverse transcriptase–polymerase chain reaction

Cells were prepared as above apart from monocytes that were positively selected over a magnetic-activated cell separation column according to the manufacturer (Miltenyi Biotech). The $CD11c^{+ve}$ and $CD11c^{-ve}$ blood DCs selected Vantage fluorescent cell sorting. Total RNA was prepared from 10 000 cells using TRIzol (Gibco) as per the manufacturer. The cDNA was synthesized from DNaseI-treated RNA with oligo-dT primers and Superscript II (Gibco). From 40 μL of RNase H–treated cDNA, 1 μL was polymerase chain reaction (PCR)–amplified with Taq polymerase (Qiagen, Germany) using either the GAPDH primers, 5′-ATGGGGAAGGTGAAGGTCGGA-3′ and 5′-AGGGGCCATCCACAGTCTTCTG-3′, to ensure equivalent amounts of cDNA in each cell type or using the first-round DC-SIGN primers, 5′AGAGTGGGGTGACATGAGTG-3′ and 5′-GAAGTTCTGCTACGCAGGAG-3′, which yielded a fragment approximately 1.2 kilobases in size. A seminested round of PCR was performed for DC-SIGN using the former 5′ primer and 5′-AGCTCCTGGTAGATCTCCTGC-3′. Electrophoresed products were transferred from a 1% agarose gel to Hybond N^+ and probed with digoxigenin-labeled internal oligonucleotide 5′-CCAGAGAAATCTAAGCTGCAGG-3′ as per the manufacturer (Roche Biochemicals, Basel, Switzerland).

HIV gp120 internalization and tracking

To examine gp120 internalization, cells were labeled with b-gp120 as described above, washed, and subsequently incubated at 37°C. For short-term incubations (< 2 hours) cells were incubated in a 37°C water, and for longer incubations (> 2 hours) cells were replated and cultured at 37°C in a 5% CO_2 incubator. Aliquots were removed at the times outlined in "Results" and terminated by incubation in 0.25% (wt/vol) paraformaldehyde in phosphate-buffered saline at 4°C for 30 minutes. For internal staining, cells were permeabilized with 0.2% (vol/vol) Tween 20, 1% (vol/vol) fetal calf serum in phosphate-buffered saline for 15 minutes at 37°C. Detection of external or internal gp120 was via streptavidin Oregon Green 488 as described above.

Results

HIV gp120 binding to CLR and/or CD4 on immature MDDCs

Because of its potent inhibition of CLRs[15] and lack of interference with gp120-CD4 binding, mannan was chosen as an inhibitory ligand to determine the proportion of gp120 bound to CLRs in MDDCs.[15] In MDDCs, mannan inhibited gp120 by up to 84% (Figure 1A). Higher levels of mannan were also used (up to 25 mg/mL), but further gp120 blocking was not observed (data not shown). Nonbiotinylated Chinese hamster ovary cell–expressed gp120 (detected via anti-HIV polyclonal antibodies) from the primary R5 isolate MW959 was also inhibited with mannan by up to 80% (data not shown). The other CLR inhibitor, α-methyl-mannopyranoside, and the calcium chelator, ethyleneglycotetraacetic acid (EGTA), inhibited gp120 binding by 82% and 77%, respectively (Figure 2). The residual gp120 binding was initially attributed to CD4. Therefore, the gp120-blocking CD4 mAbs Leu3a and Q4120, with the nonblocking mAb OKT4 as a negative control, were used to determine CD4 binding. However, neither Leu3a nor Q4120 could block gp120 binding at concentrations up to 25 μg/mL (Figure 1B). In view of this CLR-gp120 binding predominance, incubation with CD4 mAbs after prior blocking of CLR binding was examined. To achieve this, MDDCs were preincubated with 5 mg/mL mannan and then with increasing amounts of the Leu3a. In the absence of CLR binding, anti-Leu3a was successful at inhibiting the residual 10% to 20% gp120 binding to less than 1% of gp120 binding (Figure 1B).

Inhibition of mAb binding to specific CLRs by gp120

Candidate CLRs on MDDCs and other DCs for gp120 binding were DC-SIGN and MR.[14,30] Therefore, mAb DC-SIGN (AZN-D2)[14,16] and

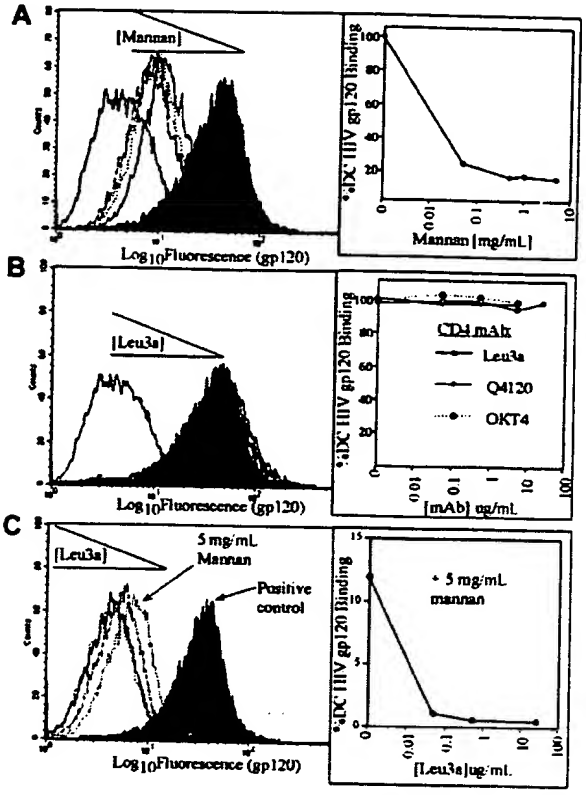


Figure 1. Inhibition of gp120 binding on MDDC. (A) Inhibition of gp120 binding to MDDCs by mannan; 1×10^6 cells/mL were incubated with mannan ranging from 50 μg to 5 mg/mL for 30 minutes at 4°C. The b-gp120 was added at 3-fold excess (9 μg/mL) and incubated for 30 minutes at 4°C. The b-gp120 was detected via streptavidin Oregon Green 488 and fluorescence measured by flow cytometry as described. (B,C) Inhibition of gp120 binding to MDDCs by CD4 mAbs (Leu3a, Q4120, OKT4). In panel B, cells were preincubated with mAbs to CD4 (ranging from 0.05 μg/mL to 25 μg/mL) for 30 minutes at 4°C. In panel C, cells were also incubated with 5 mg/mL mannan in addition to the CD4 mAb Leu3a. The b-gp120 was incubated and detected as in panel A. Percent DC-gp120 binding was calculated as follows: [(sample fluorescence intensity − mean negative control fluorescence intensity)/mean positive control fluorescence intensity] × 100. Positive control cells were treated with b-gp120 in the absence of inhibitors. Negative controls consisted of cells with identical inhibitors but no b-gp120.

MR (clone 19)[31] were used because they have been shown previously to block ligand binding. Preincubation of MDDCs with gp120 inhibited DC-SIGN (AZN-D2), MR (clone 19), and CD4 (Leu3a) mAbs in a dose-dependent manner (Figure 3A). As gp120 approached cellular saturation, binding of the mAbs to all 3 receptors approached zero. The gp120 concentrations that inhibited mAb binding by 50% (K_i) mAb were, for DC-SIGN (AZN-D2), 1 nM; MR (clone 19), 4 nM; and CD4 (Leu3a), 14 nM. The approximate dissociation constant (K_d) for BaL gp120 from the gp120 saturation curve is 6 nM for 1×10^6/mL MDDCs.

The role of individual CLRs in binding gp120

In reciprocal experiments, the effects of prior incubation with MR (clones 19 and 3.29) and DC-SIGN (AZN-D1 and AZN-D2) blocking mAbs on gp120 binding[14,16,31,32] were examined to determine relative importance of DC-SIGN and MR in gp120 binding. However, anti-MR (clones 19 and 3.29) and anti–DC-SIGN (AZN-D1 and AZN-D2) mAbs could not inhibit gp120 binding (at levels up to 5 μg/mL). The antibody bound was confirmed in each assay by goat antimouse PE, and it was confirmed that gp120 and the blocking antibodies were each bound to saturating levels on the entire MDDC population (data not

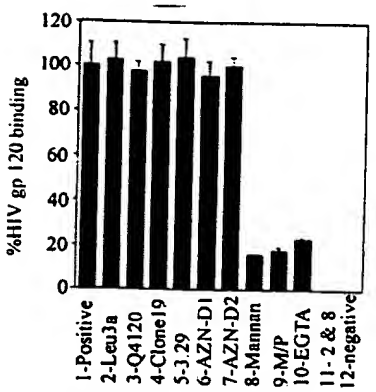


Figure 2. Inhibition of gp120 binding to MDDCs with a range of ligands. The mAbs to CD4 2-Leu3a and 3-Q4120), MR (4-clones 19 and 5-3.29), and DC-SIGN (AZN-5-D1 and 6-AZN-D2) were preincubated 5-fold above predetermined saturating concentrations (5 μg/mL). Inhibitors mannan (8-mannan), α-methyl-mannopyranoside (9-M/P), and EGTA (10-EGTA) were incubated in excess at 5 mg/mL, 125 mM, and 5mM, respectively. Dual Leu3a and mannan inhibition (11-2 and 8) included Leu3a and mannan at levels used above for treatments 2 and 8. Positive and negative controls (treatments 1 and 12) consisted of MDDCs incubated with or without gp120, respectively. The b-gp120 was added and detected as in the legend to Figure 1A.

shown). However, in the same assay mannan successfully reduced gp120 binding below 20% and combined mannan and Leu3a to below 1%. Because gp120 was used in excess in the above experiments, further inhibitory studies with 5-fold saturating concentrations of mAb (5 μg/mL) were carried out over a range of gp120 concentrations (20 ng/mL to 5 μg/mL) to observe the effects of MR and DC-SIGN mAbs (Figure 3B). However, no significant inhibition of gp120 binding by MR and/or DC-SIGN antibodies was observed at any concentration.

HIV gp120 binding to trypsin-insensitive CLRs

To address the possibility that MDDCs express several CLRs capable of binding gp120, cells were trypsinized to denude them of both the CD4-gp120 binding site and the carbohydrate recognition

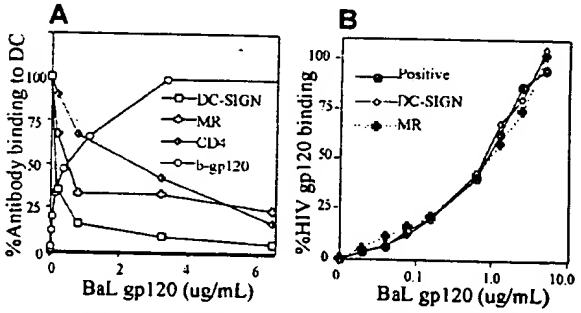


Figure 3. Interaction between gp120 and mAbs to CD4, DC-SIGN, and MR. (A) Inhibition of mAbs with increasing concentrations of gp120. MDDCs were incubated with increasing concentrations of b-gp120 under conditions outlined in Figure 1A. The availability of CD4 and CLR epitopes (those not blocked by gp120 binding) was detected by mAbs to CD4 (Leu3a), DC-SIGN (AZN-D2), and MR (clone 19) all at 1 μg/mL. For comparison, binding of b-gp120 alone at increasing concentrations is shown. Detection and incubation of bound b-gp120 was performed as outlined in Figure 1A. Percent binding of mAbs to DCs was calculated as per Figure 1 with positive and negative controls defined as follows. Positive controls were cells incubated with mAbs in the absence of gp120. Negative controls were cells incubated with the appropriate mAb isotype control (IgG_1 for all 3 mAbs listed above). (B) Inhibition of gp120 binding to CLRs by mannan, DC-SIGN (AZN-D2), and MR (clone 19) mAbs at various concentrations of gp120. MDDCs were preincubated with mAbs as in Figure 2. After washing, b-gp120 was incubated with cells at concentrations ranging from 20 ng/mL to 5 μg/mL and detected as outlined in Figure 1A.

domains (CRDs) of either DC-SIGN and/or the MR. As expected, the CD4 Leu3a epitope was cleaved. The CRD for DC-SIGN was also trypsin-sensitive, whereas the MR clone 19 epitope was not (Figure 4Biii). When trypsinized MDDCs were exposed to gp120, they retained the ability to bind gp120 at a reduced level (Figure 4Di). If trypsinized cells were preexposed to mannan or EGTA, they lost their ability to bind to gp120, indicating binding was carbohydrate- and calcium-dependent, characteristic of a trypsin-resistant CLR but clearly not DC-SIGN (Figure 4Dii, iii, respectively). To address whether this CLR might be MR, the anti-MR mAb clones 19 and 3.29 were used to block the trypsin-insensitive gp120 binding. However, MR mAb clones 19 and 3.29 could not significantly reduce trypsin-insensitive gp120 binding (Figure 4Div). To ensure that the mAbs can block gp120 binding, parallel studies were carried out with a transfected cell line expressing macrophage mannose receptor (MMR).[33] The MMR mAbs could inhibit gp120 binding to 50% regardless of whether these cells were trypsinized (data not shown); ie, the mAbs were partial inhibitors of gp120 binding to MMR on the transfected cell line but had no effect on MDDCs regardless of whether they were trypsinized.

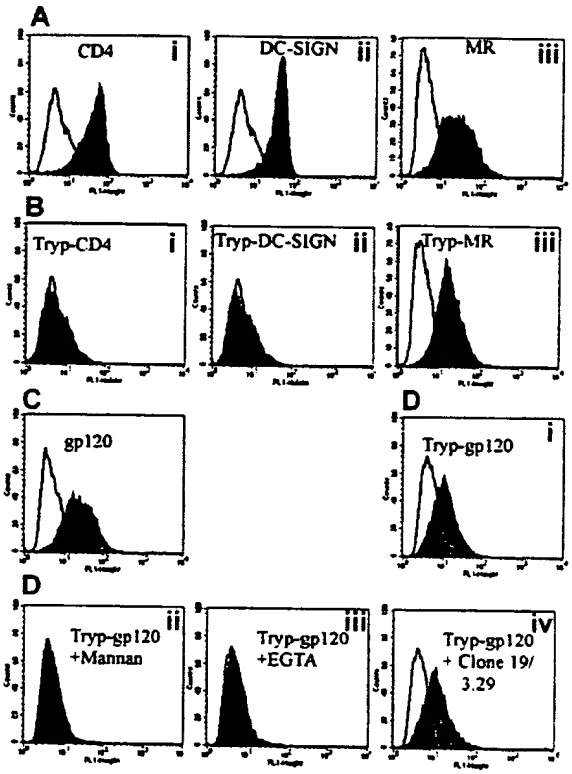


Figure 4. Effect of trypsin on CD4, MR, and DC-SIGN mAbs and b-gp120 binding to MDDCs. (A) CD4 (Leu3a) (Ai), DC-SIGN (AZN-D2) (Aii), and MR (clone 19) (Aiii) staining before trypsinization. A total of 2 μg/mL of mAb to CD4, DC-SIGN, and MR was added as outlined in "Materials and methods." The mAb binding was detected via goat antimouse FITC (1 μg/mL) (Becton Dickinson) and fluorescence measured as in Figure 1. Gray histograms represent antibody staining with open overlaid histogram staining by matching isotype controls. (B) CD4 (Leu3a) (Bi), DC-SIGN (AZN-D2) (Bii), and MR (clone 19) (Biii) staining after trypsinization. Cells were treated with 0.25% trypsin at 37°C for 5 minutes and subsequently washed in normal media before the addition of mAbs to CD4, DC-SIGN, and MR as in panel A. (C) The b-gp120 binding before trypsinization. (D) The b-gp120 binding to MDDCs after trypsinization: effect of inhibitors. Trypsinized cells were mock-treated (Di) or treated with excess mannan (5 mg/mL) (Dii), EGTA (5 mM) (Diii), or anti-MR (clones 19 and 3.29) (5 μg/mL) (Div) for 30 minutes at 4°C. The b-gp120 was added and detected as in Figure 1A. Gray histograms represent gp120 staining and open overlays matched negative controls (treatment without addition of b-gp120).

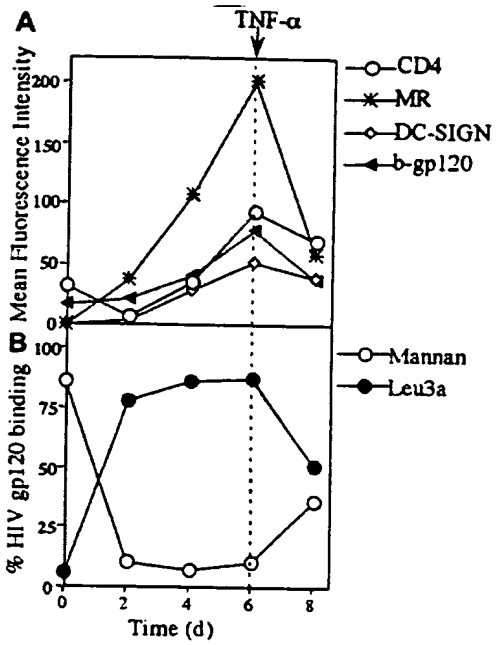


Figure 5. Kinetics of CD4, DC-SIGN, and MR expression and gp120 binding during differentiation of monocytes to immature and mature MDDCs. (A) CD4 and CLR expression and b-gp120 binding. Monocytes were stimulated to immature MDDCs as described in "Materials and methods." Mature MDDCs were generated from day 6 to 8 by addition of 10 ng/mL TNF-α and expressed both maturation markers CD83 and CD86 by day 8 (> 70% +ve for both markers). The b-gp120, MAb to CD4 (leu3a), DC-SIGN (AZN-D2), and MR (clone 19) were added, incubated, and detected at days 0, 2, 4, 6, and 8 as outlined in Figures 1A and 4. The mean relative intensity of the isotype or negative control was subtracted from the mean fluorescent intensity for 10 000 cells. (B) HIV gp120 binding to CD4 and CLRs. At days 0, 2, 4, 6, and 8, cells were preincubated with either saturating levels of Leu3a (10 μg/mL) or mannan (5 mg/mL) and subsequently exposed to saturating levels of b-gp120 as outlined in Figures 1 and 2.

HIV gp120 binding during differentiation of monocytes to MDDCs

The switch from gp120 binding to CD4 on monocytes to CLRs on MDDCs was examined during in vitro differentiation over 6 days. By day 2, CLR binding was predominant (Figure 5B) and correlated with a rise in MR expression and CD4 down-regulation (Figure 5A). Over day 2 to day 6 of differentiation, there was a continuous increase in binding of gp120 to CLR with a corresponding decrease in CD4 binding. Over the same period, there was a continuous increase in DC-SIGN, CD4, and MR expression. The peak expression of all 3 receptors at day 6 coincided with the peak in gp120 binding (Figure 5A). Mature MDDCs were generated by stimulation with TNF-α for 2 days. After maturation, MR, DC-SIGN, and CD4 were all down-regulated, but this was more marked with MR (Figure 5A). In mature MDDCs, the pattern of gp120 binding to CLRs and CD4 converged, with intermediate levels of binding to both (Figures 5B and 7B).

HIV gp120 binding on ex vivo blood DCs

Because MDDCs are derived in vitro, it was important to determine the gp120 binding receptors on ex vivo blood DCs. Blood DCs were separated, incubated with gp120, and triple-stained for b-gp120, CD11c, and HLA-DR, which allowed identification of 2 blood DC populations based on the presence or absence of CD11c

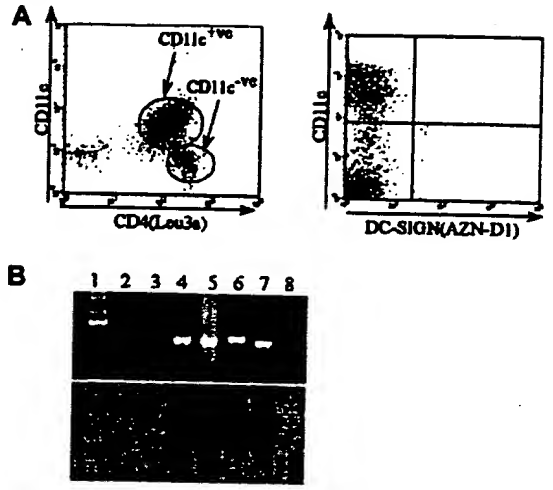


Figure 6. CD4 and DC-SIGN expression on blood DC. (A) CD4 and DC-SIGN surface expression on blood DC subsets. Blood DCs were freshly isolated as outlined in "Materials and methods." The mAb to CD4 (Leu3a) or DC-SIGN (AZN-D1) was added and then incubated and detected as outlined in Figure 4. Cultured DCs were DC-incubated overnight in the presence of interleukin-3/GM-CSF. Blood DC subsets were further distinguished by CD11c staining. (B) DC-SIGN expression on blood DCs by RT-PCR. Top panel: ethidium bromide–stained gel (top) of PCR products for DC-SIGN from cDNA. Lane 1: 1-kilobase ladder; lane 2: CD11c^{+ve} blood DCs; lane 3: CD11c^{-ve} blood DCs; lane 4: monocytes; lane 5: MDDCs; lane 6: MDDCs cultured with lipopolysaccharide; Lane 7: PBMCs; lane 8: H_2O. Bottom panel: the autoradiograph of the Southern blot probed with a digoxigenin oligonucleotide specific for DC-SIGN.

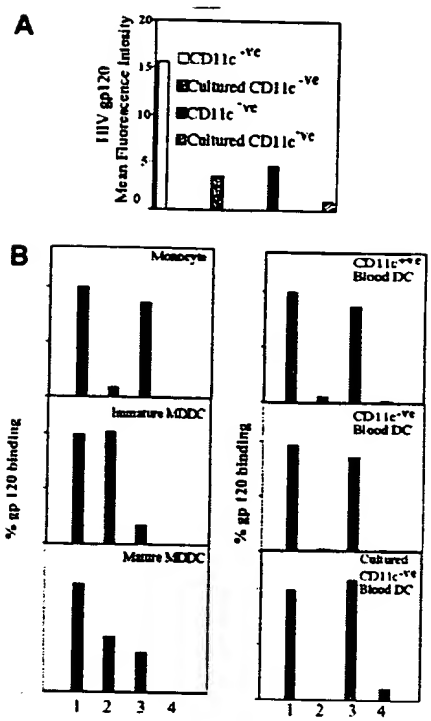


Figure 7. Binding of gp120 on several DC subsets. (A) Relative gp120 binding levels within blood DC subsets. Blood DCs were freshly isolated as outlined in "Materials and methods." The b-gp120 (Leu3a) was added and then incubated and detected as outlined in Figure 1A. Cultured blood DCs were incubated as outlined in Figure 6. Blood DC subsets were further analyzed by CD11c staining after b-gp120 staining. (B) Inhibition of gp120 binding by anti-CD4 (Leu3a) and mannan to MDDCs and blood DCs. Inhibitors were preincubated with blood DCs as follows: (1) No inhibitors, (2) CD4 (Leu3a) (10 μg/mL), (3) mannan (5 mg/mL), and (4) dual mannan/Leu3a were preincubated with blood DCs. The b-gp120 was added as outlined in Figure 1A and DC subsets analyzed for gp120 binding after CD11c staining.

expression. The CD11c^{-ve} population expressed much higher levels of CD4 (Figure 6A) and bound greater amounts of gp120 than the CD11c^{+ve} population (Figure 7A). CD4 was down-regulated on both blood DC subsets after overnight culture (data not shown) and was reflected by the reduced capacity to bind gp120 (Figure 7A). The importance of CLRs and CD4 for gp120 binding was determined by blocking experiments with mannan and anti-CD4 (Leu3a) mAbs (Figure 7B). The pattern of binding was similar on both blood DC subsets—both fresh and after overnight culture—with a predominance of gp120 binding to CD4 rather than CLRs. The lack of CLR binding was supported by the lack of MR (data not shown) and DC-SIGN surface expression (Figure 6A). Semi-nested reverse transcriptase (RT)–PCR for DC-SIGN confirmed lack of messenger RNA transcripts in both blood DC subsets. However, transcripts were seen in PBMC and CD14^{+ve} monocyte populations (Figure 6B).

HIV gp120 internalization

Internalization was rapid, with less than 50% of surface gp120 present after 5 minutes. After 1 hour no external gp120 could be observed on MDDCs (Figure 8). MDDCs were reexamined for surface gp120 over 2, 6, 18, and 24 hours. There was no reappearance of external gp120 over the period of 1 to 24 hours. The kinetics of gp120 internalization mediated by CD4 and CLRs was also investigated. First, the CLR pathway was blocked by mannan, and gp120 bound to CD4 was examined for internalization. Conversely, the role of CLRs in internalization was also examined by blocking CD4 with Leu3a and gp120 subsequently tracked. Both CD4 and CLR pathways exhibited rapid internalization with no external gp120 evident after 60 minutes. The CD4-mediated internalization pathway showed a single rapid phase, but CLR internalization was biphasic. The first phase rapidly internalized most of the gp120 within the first 15 minutes, and the second phase internalized the residual gp120 over the 15- to 60-minute period. Rapid external loss of gp120 correlated with rapid appearance of internalized gp120 as observed in permeabilized MDDCs (Figure 8).

Discussion

MDDCs were used in the current studies as a model for immature tissue DCs such as skin LCs and mucosal DCs. They are a

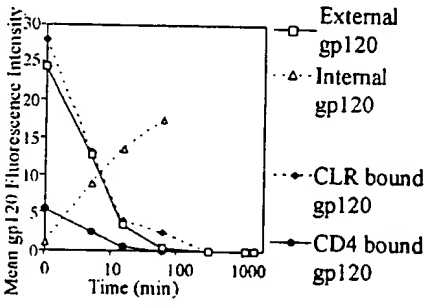


Figure 8. Internalization of gp120 by MDDCs. Cells were preincubated with anti-CD4 (Leu3a) to detect CLR-bound gp120 and with mannan to detect CD4-bound gp120 or binding media for total external or internal gp120 for 30 minutes at 4°C. Cells were then incubated with 5 μg/mL b-gp120 for 30 minutes at 4°C, washed twice in binding media, and incubated in culture media for the indicated times and stained to detect extracellular or intracellular gp120 as outlined in "Materials and methods."

convenient model for in vitro studies but also may have relevance in vivo: Monocytes are observed to develop into MDDCs at sites of inflammation as a second recruitment of antigen-presenting cells.[34] Nethertheless, they show marked phenotypic difference to other blood and tissue DCs.[35] Therefore, we defined the receptors for binding gp120 on MDDCs in vitro and then compared them with ex vivo blood DCs.

In MDDCs, 2 groups of receptors capable of binding gp120 were defined. MDDCs bound gp120 predominantly via CLRs: the mannose saccharides, mannan and mannopyranoside, and also calcium depletion were capable of markedly inhibiting gp120 binding. Monocytes only bound gp120 via CD4 and did not express MR or DC-SIGN. Conversion to the predominant CLR binding pattern seen in MDDCs occurred on monocytes after 2 days of culture in interleukin-4/GM-CSF and peaked at day 6. During MDDC differentiation, the kinetics of DC-SIGN, MR, and CD4 expression and gp120 binding via CLRs were discordant, which supports a more complex gp120 binding pattern than previously described. TNF-α–induced MDDC maturation increased CD4-gp120 binding at the expense of CLR binding and also significantly reduced MR but only slightly decreased CD4 and DC-SIGN expression.

Both $CD11c^{+ve}$ and $CD11c^{-ve}$ blood DCs lacked both DC-SIGN and MR expression, and gp120 bound exclusively by CD4. Culture of both blood DC subsets down-regulated CD4 expression and gp120 binding but did not induce MR, DC-SIGN expression, or gp120 binding via CLRs.

The 2 CLRs, DC-SIGN and the MR, have been previously observed to bind gp120,[14,15,30] and both are expressed on MDDCs. HIV gp120 bound to the surface of MDDCs and inhibited anti-CD4, anti-MR, and anti-DC-SIGN mAb binding, supporting gp120 binding to the above 3 receptors. DC-SIGN mAb was most readily inhibited at low gp120 concentrations, consistent with high affinity for gp120.[15] However, neither CD4, DC-SIGN, nor MR mAb inhibited gp120 binding to MDDCs. Trypsin treatment of MDDCs completely cleaved both the CD4 (Leu3a) and DC-SIGN (AZN-D2) mAb epitopes but only partially inhibited gp120 binding to MDDCs. Both MR mAb clones 19 and 3.29 still bound to trypsinized MDDCs, probably to CRDs 4 or 5, which are protease-insensitive.[33] Residual gp120 binding, in trypsinized MDDCs, was blocked by mannan and EGTA but not by either MR mAbs. These results suggest gp120 could bind to other CLRs and/or other CRDs of MR (not recognized by the mAbs). However, the latter seems unlikely. because both mAbs block binding of mannose ligands to MR[31,32] and, more specifically, partially block gp120 in a trypsinized MMR cell line (data not shown). If several CLRs, including DC-SIGN and MR, can bind gp120, blocking one CLR with mAbs may not significantly reduce gp120 binding. This notion is further supported by the inability of either CD4 or CLR mAbs alone to inhibit binding. In addition, the binding of gp120 to CD4 differed in the presence or absence (mannan block) of CLRs. This might reflect the much higher binding affinity of the CLRs (MMR and DC-SIGN, $K_d < 4$ nM) compared with the CD4 affinity for BaL gp120 ($K_d = 30$ nM).

Experiments on gp120 internalization independently confirmed that gp120 bound predominantly via CLRs. The rapid internalization of gp120 in COS-7–DC-SIGN transfectants observed by Curtis et al[15] and in HeLa transfectants (A. J. Watson, written communication, August 2000), together with internalization of the MR,[36] supports our observation of a rapid CLR-mediated phase of gp120 internalization. The biphasic nature of this CLR-based internalization could reflect multiple CLRs capable of binding and internalizing gp120. Electron microscopic studies by Blauvelt et al,[8] Dezutter-Dambuyant and Schmitt,[37] and Hladik et al[7] showed internalization of virions into vacuoles and is consistent with current observations of gp120 internalization. In electron microscopy studies by Dezutter-Dambuyant and Schmitt,[37] HIV gp120 internalization was correlated with whole virions, because both were observed in clathrin-coated pits of epidermal LCs. Similarly stable HeLa clone 11 (DC-SIGN) transfectants also internalized HIV into vacuoles, suggesting that CLR binding results in endocytosis (A. J. Watson, written communication, August 2000). In our recent work, mannan was also shown to markedly inhibit accumulation of full-length HIV proviral DNA transcripts within MDDCs, showing a close correlation between gp120 internalization and HIV infection (unpublished observations, 2001). In the current study there was no reappearance of gp120 on the surface of MDDCs, suggesting there was degradation after internalization.

There are many reports of the ability of gp120 to bind to various cell types independently of CD4. Macrophages,[30] trypsinized LCs,[13,38] MDDCs,[16] and cells within the placenta[15] are examples. Only the studies of Curtis et al[15] and Larkin et al[30] identified the specific receptors as CLRs. Geijtenbeek et al[14,16] recently reported that placental CLR clone 11 (DC-SIGN) previously described by Curtis et al[15] was expressed on MDDCs. While the observations described here support CLRs as predominant receptors for gp120 binding to MDDCs, CLR binding of gp120 was not restricted to one receptor as reported previously[14] but instead to multiple CLRs, including DC-SIGN and MR. A further CLR related to DC-SIGN, named DC-SIGNR, has recently been indentified on MDDCs,[21] and the potential expression and binding by numerous other CLRs on MDDCs[17-20] further supports our current hypothesis that multiple CLRs can bind gp120.

Although CLRs bound most gp120 in MDDCs, CD4 is the predominant receptor in blood DCs. This observation expands previously described phenotypic differences between MDDCs and blood DCs.[35] Thus, the fate of internalized gp120 or of HIV is highly likely to be determined by initial binding to CLRs (MDDCs) or CD4 and then the appropriate chemokine receptors (blood DCs). Transfer of HIV from blood DCs to T cells as shown by Cameron et al[4] must involve initial binding by CD4. In contrast, Blauvelt et al[8] observed that in vitro–derived DCs have the capacity to capture and transfer HIV independently of the CD4/chemokine receptor infection pathway. The current work and recent work by Geijtenbeek et al[14] suggest that this previously unknown capture pathway is by CLRs. However, both MDDCs and blood DCs capture and transfer HIV to CD4 T lymphocytes effectively in coculture assays. In light of the current observations, it is obvious that blood DCs could not capture and transfer HIV via both pathways. Further viral binding mechanisms independent of CD4 and CLR may also be present. For instance, HIV can acquire T cell–specific molecules during budding,[39,40] and DCs may be able to bind virions via the same mechanism they use in clustering to T cells. Another DC in vivo, the follicular DC, predominantly binds HIV virions via the adhesion molecules CD54 (ICAM-1) and CD11a (LFA-1).[41] Macropinocytosis must also be considered as another mechanism of gp120/viral uptake by DCs.

In view of the discordant findings for gp120 binding between MDDCs and blood DCs, future work must focus on which CLRs are expressed in vivo on LCs and mucosal DCs and whether CLRs or CD4 are the major receptors for gp120 in these cells. LCs do not express DC-SIGN[16] and expression of the MR is controversial,[24,42] but they do express a mannose-fucose binding receptor(s)[42] and can bind gp120 independently of CD4.[13] Therefore, other CLRs and/or

CD4/CCR5 could be even more important than DC-SIGN in studies of DC-mediated HIV mucosal transmission. The study of the relevant receptors in appropriate surveillance DCs is essential to understanding both mucosal HIV transmission and systemic or mucosal gp120 antigenic processing pathways. These results are relevant to the design of effective antivirals: Care must be taken to ensure that all routes of HIV-DC binding are blocked, because DCs may bind and transfer HIV to responding CD4 T cells via several of their cell surface receptors.

Acknowl dgments

We thank Ray Sweet for his generous gift of BaL gp120, Yvette van Kooyk for the supply of both antibodies AZN-D1 and AZN-D2 to DC-SIGN, Belinda Herring for her help and assistance, Parramatta Red Cross Blood Bank for its continued support, and Nancy Haigwood and Andrew Watson for helpful discussions on the placental clone 11.

References

1. Edwards JN, Morris HB. Langerhans' cells and lymphocyte subsets in the female genital tract. Br J Obstet Gynaecol. 1985;92:974-982.
2. Miller CJ, McChesney M, Moore PF. Langerhans cells, macrophages and lymphocyte subsets in the cervix and vagina of rhesus macaques. Lab Invest. 1992;67:628-634.
3. Hu J, Gardner MB, Miller CJ. Simian immunodeficiency virus rapidly penetrates the cervicovaginal mucosa after intravaginal inoculation and infects intraepithelial dendritic cells. J Virol. 2000;74: 6087-6095.
4. Cameron PU, Freudenthal PS, Barker JM, Gezelter S, Inaba K, Steinman RM. Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4+ T cells. Science. 1992;257:383-387.
5. Pope M, Betjes MG, Romani N, et al. Conjugates of dendritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1. Cell. 1994;78:389-398.
6. Pope M, Elmore D, Ho D, Marx P. Dendrite cell-T cell mixtures, isolated from the skin and mucosae of macaques, support the replication of SIV. AIDS Res Hum Retroviruses. 1997;13:819-827.
7. Hladik F, Lentz G, Akridge RE, et al. Dendritic cell-T-cell interactions support coreceptor-independent human immunodeficiency virus type 1 transmission in the human genital tract. J Virol. 1999;73:5833-5842.
8. Blauvelt A, Asada H, Saville MW, et al. Productive infection of dendritic cells by HIV-1 and their ability to capture virus are mediated through separate pathways. J Clin Invest. 1997;100:2043-2053.
9. Patterson S, Robinson SP, English NR, Knight SC. Subpopulations of peripheral blood dendritic cells show differential susceptibility to infection with a lymphotropic strain of HIV-1. Immunol Lett. 1999;66:111-116.
10. Rubbert A, Combadiere C, Ostrowski M, et al. Dendritic cells express multiple chemokine receptors used as coreceptors for HIV entry. J Immunol. 1998;160:3933-3941.
11. Granelli-Piperno A, Delgado E, Finkel V, Paxton W, Steinman RM. Immature dendritic cells selectively replicate macrophagetropic (M-tropic) human immunodeficiency virus type 1, while mature cells efficiently transmit both M- and T-tropic virus to T cells. J Virol. 1998;72:2733-2737.
12. Zaitseva M, Blauvelt A, Lee S, et al. Expression and function of CCR5 and CXCR4 on human Langerhans cells and macrophages: implications for HIV primary infection. Nat Med. 1997;3:1369-1375.
13. Schmitt D, Dezutter-Dambuyant C, Hanau D, et al. In vitro binding and internalization of HIV envelope glycoproteins by human epidermal Langerhans cells does not require the CD4-gp120-binding site. Res Virol. 1990;141:209-215.
14. Geijtenbeek TB, Kwon DS, Torensma R, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell. 2000;100:587-597.
15. Curtis BM, Scharnowske S, Watson AJ. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proc Natl Acad Sci U S A. 1992;89:8356-8360.
16. Geijtenbeek TB, Torensma R, van Vliet SJ, et al. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell. 2000;100:575-585.
17. Ariizumi K, Shen GL, Shikano S, et al. Cloning of a second dendritic cell-associated C-type lectin (dectin-2) and its alternatively spliced isoforms. J Biol Chem. 2000;275:11957-11963.
18. Bakker AB, Baker E, Sutherland GR, Phillips JH, Lanier LL. Myeloid DAP12-associating lectin (MDL)-1 is a cell surface receptor involved in the activation of myeloid cells. Proc Natl Acad Sci U S A. 1999;96:9792-9796.
19. Bates EE, Fournier N, Garcia E, et al. APCs express DCIR, a novel C-type lectin surface receptor containing an immunoreceptor tyrosine-based inhibitory motif. J Immunol. 1999;163:1973-1983.
20. Colonna M, Samaridis J, Angman L. Molecular characterization of two novel C-type lectin-like receptors, one of which is selectively expressed in human dendritic cells. Eur J Immunol. 2000;30: 697-704.
21. Soilleux EJ, Barten R, Trowsdale J. DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. J Immunol. 2000;165:2937-2942.
22. Stahl PD, Ezekowitz RA. The mannose receptor is a pattern recognition receptor involved in host defense. Curr Opin Immunol. 1998;10:50-55.
23. Kato M, Neil TK, Clark GJ, Morris CM, Sorg RV, Hart DNJ. cDNA cloning of human DEC-205, a putative antigen-uptake receptor on dendritic cells. Immunogenetics. 1998;47:442-450.
24. Kato M, Neil TK, Fearnley DB, McLellan AD, Vuckovic S, Hart DN. Expression of multilectin receptors and comparative FITC-dextran uptake by human dendritic cells. Int Immunol. 2000;12: 1511-1519.
25. Mason RR, Weiner RS. Application of the Beckman JE6-B Elutriator System in the isolation of human monocyte subpopulations. Scand J Haematol. 1985;34:5-8.
26. Kazazi F, Mathijs JM, Chang J, et al. Recombinant interleukin 4 stimulates human immunodeficiency virus production by infected monocytes and macrophages. J Gen Virol. 1992;73:941-949.
27. Romani N, Gruner S, Brang D, et al. Proliferating dendritic cell progenitors in human blood. J Exp Med. 1994;180:83-93.
28. Sallusto F, Lanzavecchia A. Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. J Exp Med. 1994;179:1109-1118.
29. Fear WR, Kesson AM, Naif H, Lynch GW, Cunningham AL. Differential tropism and chemokine receptor expression of human immunodeficiency virus type 1 in neonatal monocytes, monocyte-derived macrophages, and placental macrophages. J Virol. 1998;72:1334-1344.
30. Larkin M, Childs RA, Matthews TJ, et al. Oligosaccharide-mediated interactions of the envelope glycoprotein gp120 of HIV-1 that are independent of CD4 recognition. AIDS. 1989;3:793-798.
31. Dong X, Storkus WJ, Salter RD. Binding and uptake of agalactosyl IgG by mannose receptor on macrophages and dendritic cells. J Immunol. 1999;163:5427-5434.
32. Sallusto F, Cella M, Danieli C, Lanzavecchia A. Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. J Exp Med. 1995; 182:389-400.
33. Taylor ME, Drickamer K. Structural requirements for high affinity binding of complex ligands by the macrophage mannose receptor. J Biol Chem. 1993;268:399-404.
34. Randolph GJ, Inaba K, Robbiani DF, Steinman RM, Muller WA. Differentiation of phagocytic monocytes into lymph node dendritic cells in vivo. Immunity. 1999;11:753-761.
35. Hart DN. Dendritic cells: unique leukocyte populations which control the primary immune response. Blood. 1997;90:3245-3287.
36. Tan MC, Mommaas AM, Drijfhout JW, et al. Mannose receptor mediated uptake of antigens strongly enhances HLA-class II restricted antigen presentation by cultured dendritic cells. Adv Exp Med Biol. 1997;417:171-174.
37. Dezutter-Dambuyant C, Schmitt D. Epidermal Langerhans cells and HIV-1 infection. Immunol Lett. 1993;39:33-37.
38. Dezutter-Dambuyant C, Schmitt DA, Dusserre N, et al. Trypsin-resistant gp120 receptors are upregulated on short-term cultured human epidermal Langerhans cells. Res Virol. 1991;142:129-138.
39. Liao Z, Roos JW, Hildreth JE. Increased infectivity of HIV type 1 particles bound to cell surface and solid-phase ICAM-1 and VCAM-1 through acquired adhesion molecules LFA-1 and VLA-4. AIDS Res Hum Retroviruses. 2000;16:355-366.
40. Frank I, Kacani L, Stoiber H, et al. Human immunodeficiency virus type 1 derived from cocultures of immature dendritic cells with autologous T cells carries T-cell-specific molecules on its surface and is highly infectious. J Virol. 1999;73:3449-3454.
41. Fujiwara M, Tsunoda R, Shigeta S, Yokota T, Baba M. Human follicular dendritic cells remain uninfected and capture human immunodeficiency virus type 1 through CD54-CD11a interaction. J Virol. 1999;73:3603-3607.
42. Condaminet B, Peguet-Navarro J, Stahl PD, Dalbiez-Gauthier C, Schmitt D, Berthier-Vergnes O. Human epidermal Langerhans cells express the mannose-fucose binding receptor. Eur J Immunol. 1998;28:3541-3551.

A Dendritic Cell–specific Intercellular Adhesion Molecule 3–grabbing Nonintegrin (DC-SIGN)–related Protein Is Highly Expressed on Human Liver Sinusoidal Endothelial Cells and Promotes HIV-1 Infection

By Arman A. Bashirova,* Teunis B.H. Geijtenbeek,‖ Gerard C.F. van Duijnhoven,‖ Sandra J. van Vliet,‖ Jeroen B.G. Eilering,‖ Maureen P. Martin,‡ Li Wu,§ Thomas D. Martin,§ Nicola Viebig,¶ Percy A. Knolle,¶ Vineet N. KewalRamani,§ Yvette van Kooyk,‖ and Mary Carrington‡

*From the *Laboratory of Genomic Diversity, the ‡Intramural Research Support Program, Science Applications International Corporation-Frederick, and the §HIV Drug Resistance Program, National Cancer Institute-Frederick, Frederick, Maryland 21702; the ‖Tumor Immunology Department, University Medical Center St. Radbound, Nijmegen 6525 EX, The Netherlands; and the ¶Zentrum für Moleculare Biologie Heidelberg (ZMBH), D-69120 Heidelberg, Germany*

Abstract

The discovery of dendritic cell (DC)-specific intercellular adhesion molecule (ICAM)-3–grabbing nonintegrin (DC-SIGN) as a DC-specific ICAM-3 binding receptor that enhances HIV-1 infection of T cells in trans has indicated a potentially important role for adhesion molecules in AIDS pathogenesis. A related molecule called DC-SIGNR exhibits 77% amino acid sequence identity with DC-SIGN. The *DC-SIGN* and *DC-SIGNR* genes map within a 30-kb region on chromosome 19p13.2-3. Their strong homology and close physical location indicate a recent duplication of the original gene. Messenger RNA and protein expression patterns demonstrate that the DC-SIGN–related molecule is highly expressed on liver sinusoidal cells and in the lymph node but not on DCs, in contrast to DC-SIGN. Therefore, we suggest that a more appropriate name for the DC-SIGN–related molecule is L-SIGN, liver/lymph node–specific ICAM-3–grabbing nonintegrin. We show that in the liver, L-SIGN is expressed by sinusoidal endothelial cells. Functional studies indicate that L-SIGN behaves similarly to DC-SIGN in that it has a high affinity for ICAM-3, captures HIV-1 through gp120 binding, and enhances HIV-1 infection of T cells in trans. We propose that L-SIGN may play an important role in the interaction between liver sinusoidal endothelium and trafficking lymphocytes, as well as function in the pathogenesis of HIV-1.

Key words: L-SIGN • adhesion receptor • chromosome 19p13.2-3 • ICAM-3 • HIV-1 gp120

Introduction

Dendritic cell (DC)[1]-specific intercellular adhesion molecule (ICAM)-3–grabbing nonintegrin (DC-SIGN) has recently been identified as a DC-specific adhesion receptor that mediates the interaction between DCs and resting T cells through high affinity binding to ICAM-3, thereby facilitating the initiation of primary immune responses (1, 2). DC-SIGN was shown to be identical to the previously reported type II membrane-associated C-type lectin (2) that binds HIV-1 envelope glycoprotein gp120 in a CD4-independent manner (3). The affinity of DC-SIGN exceeds that of CD4 for HIV-1 gp120 (3), and upon capture of HIV-1, DC-SIGN does not appear to promote viral entry into the DC itself, but rather enhances infection of T cells in trans (1). DC-SIGN–associated HIV-1 remains infectious over a prolonged period of time, perhaps contributing to the infectious potential of the virus during its transport by DCs from the periphery to lymphoid organs.

A.A. Bashirova and T.B.H. Geijtenbeek contributed equally to this work.

Address correspondence to Mary Carrington, P.O. Box B, NCI-FCRDC, Frederick, MD 21702. Phone: 301-846-1390; Fax: 301-846-1909; E-mail: carringt@mail.ncifcrf.gov

[1]*Abbreviations used in this paper:* CCR, CC chemokine receptor; DC, dendritic cell; DC-SIGN, DC-specific ICAM-3–grabbing nonintegrin; EST, expressed sequence tag; ICAM, intercellular adhesion molecule; L-SIGN, liver/lymph node–specific ICAM-3–grabbing nonintegrin; LSEC, liver sinusoidal endothelial cell; nt, nucleotides; RH, radiation hybrid; RT, reverse transcription; UTR, untranslated region.

A previous search by Yokoyama-Kobayashi et al. (4) for cDNA clones encoding type II membrane proteins resulted in the identification of a partial clone that was homologous, but not identical, to the cDNA encoding the molecule now known as DC-SIGN. The putative protein product contained a deletion of 28 amino acids in the cytoplasmic domain and was lacking the entire C-type lectin domain relative to the cDNA encoding DC-SIGN. More recently, Soilleux et al. (5) described the full-length cDNA sequence of the related gene, which they called *DC-SIGNR*. The genomic organization of *DC-SIGN* and *DC-SIGNR* was compared, indicating a high degree of similarity. Concomitant expression of the two genes in placenta, endometrium, and stimulated KG1 cells (a cell line that phenotypically resembles myeloid DCs) was observed, although the expression of *DC-SIGNR* was very low in both endometrium and stimulated KG1 cells (5).

While attempting to identify polymorphisms in the *DC-SIGN* gene, we also discovered the gene corresponding to the partial cDNA sequence described by Yokoyama-Kobayashi et al. (4). Tissue expression patterns of the *DC-SIGNR* gene indicated that it is expressed at considerably high levels in only two tissues, liver and lymph node, but not in monocyte-derived DCs. Therefore, we have called the molecule L-SIGN, liver/lymph node–specific ICAM-3–grabbing nonintegrin, which we believe more accurately depicts the function and expression pattern of this molecule than does DC-SIGNR. Here we refine the genomic organization of the *SIGN* gene complex, and also report the tissue distribution and functional characterization of the L-SIGN molecule.

Materials and Methods

Characterization of DC-SIGN and L-SIGN cDNA. We have submitted the full *DC-SIGN* and *L-SIGN* cDNA sequences to GenBank/EMBL/DDBJ under accession nos. AF290886 and AF290887, respectively. The *L-SIGN* cDNA sequence represents a variant containing six repeats in exon 4. The 5′ and 3′ ends of the transcripts (except the 3′ end of the *DC-SIGN* mRNA) were determined by 5′ rapid amplification of cDNA ends (RACE; CLONTECH Laboratories, Inc.). The length of the 3′ end of the *DC-SIGN* mRNA was estimated based on Northern blot analysis data (transcript size) and reverse transcription (RT)-PCR data using forward primers specific for the 1.3-kb *DC-SIGN* cDNA sequence (3) and reverse primers specific for several GenBank/EMBL/DDBJ expressed sequence tags (ESTs) (e.g., AI472111 and AA454170), mapping downstream of the alleged 3′ end of *DC-SIGN*. A cDNA fragment containing the full coding sequence of *L-SIGN* (nucleotides [nt] 39–1184, GenBank/EMBL/DDBJ accession no. AF290887) was amplified from human placental mRNA (CLONTECH Laboratories, Inc.) and cloned into the expression vectors pcDNA3.1/V5-His/TOPO (pcDNA3-L-SIGN) and pCDM8 (pCDM8-L-SIGN).

Radiation Hybrid Mapping. PCR-based radiation hybrid (RH) mapping with *DC-SIGN*– and *L-SIGN*–specific primers was performed using the Genebridge 4 RH panel (Research Genetics). The PCR results were submitted to the Gene Map server at the Sanger Center (http://www.sanger.ac.uk/Software/Rhserver). The chromosomal position of markers linked to the genes was determined searching the Genatlas database (http://web.citi2.fr/GENATLAS) and the genetic map of human chromosome 19 provided by the Marshfield Clinic (http://research.marshfieldclinic.org/genetics/).

Genotype Analysis of L-SIGN and DC-SIGN Exon4. The repeat region in exon 4 was amplified with the following pairs of primers: L28, TGTCCAAGGTCCCCAGCTCCC, and L32, GAACTCACCAAATGCAGTCTTCAAATC, for *L-SIGN*; DL27, TGTCCAAGGTCCCCAGCTCC, and DI4R, CCCCGTGTTCTCATTTCACAG, for *DC-SIGN*. The cycle conditions were as follows: 94°C for 5 s and 68°C for 1 min. Alleles were distinguished by agarose gel electrophoresis and ethidium bromide staining.

Northern Blot Analysis. Total RNA from cultured human immature DCs (see below) was isolated using Trizol (Life Technologies). 10 μg of the isolated RNA was electrophoresed on a 1% agarose gel, transferred to Hybond-XL (Amersham Pharmacia Biotech) as described (6), and used for Northern blot analysis along with two human multiple tissue Northern blots (CLONTECH Laboratories, Inc.). Three probes were subsequently hybridized to the blots: (1) an *L-SIGN*–specific probe (nt 100–183, GenBank/EMBL/DDBJ accession no. AF290887); (2) a probe recognizing both *DC-SIGN* and *L-SIGN* (nt 1–1233, GenBank/EMBL/DDBJ accession no. AF290886); and (3) an actin control probe (CLONTECH Laboratories, Inc.). Hybridization procedures were performed according to manufacturer specifications (CLONTECH Laboratories, Inc.).

Antibodies. Anti–DC-SIGN mAbs AZN-D1 and AZN-D2 were described previously (2). mAb AZN-D3 was obtained by screening hybridoma supernatants of BALB/c mice immunized with THP-1–DC-SIGN cells (1) for the ability to stain both DC-SIGN and L-SIGN. Anti–DC-SIGN mAb AZN-D2 also cross-reacts with L-SIGN, as was initially determined by the staining of K562–L-SIGN cells (data not shown). Anti–L-SIGN rabbit antiserum was generated by immunization with two L-SIGN–specific peptides, PTTSGIRLFPRD and WNDNRCDVDNYW (Veritas, Inc. Laboratories).

Cells. DCs were cultured from monocytes in the presence of 500 U/ml IL-4 and 800 U/ml GM-CSF (Schering-Plough; references 7 and 8). At day 7 the cells expressed high levels of MHC class I and II, αMβ2 (CD11b), αXβ2 (CD11c), DC-SIGN and ICAM-1, moderate levels of LFA-1 and CD86, and low levels of CD14, as measured by flow cytometry. Stable K562 transfectants expressing L-SIGN (K562–L-SIGN) were generated by cotransfection of K562 with the pCDM8–L-SIGN plasmid and the pGK-neo vector by electroporation (9). Stable K562–DC-SIGN transfectants were generated in a similar manner using pRc/CMV–DC-SIGN (2). THP-1–DC-SIGN cells were described previously (2). Stable THP-1–L-SIGN transfectants were generated by electroporation of THP-1 cells with pcDNA3–L-SIGN, selection for G418 resistance, and positive sorting for L-SIGN expression using mAb AZN-D3. All cell lines were maintained in RPMI 1640 supplemented with 10% fetal bovine serum in addition to specific cytokine or antibiotic requirements as indicated. K562 and THP-1 are monocytic cell lines. HEK293T are human embryonic kidney cells containing a single temperature-sensitive allele of SV-40 large T antigen. GHOST cells are HIV-indicator cells derived from human osteosarcoma cells (10). Hut/CC chemokine receptor (CCR)5 cells are the transformed human T cell line Hut78 stably transduced with CCR5.

Fluorescent Beads Adhesion Assay. Carboxylate-modified TransFluorSpheres (488/645 nm, 1.0 μm; Molecular Probes) were coated with ICAM-3 as described previously for ICAM-1

(11). Fluorescent beads were coated with M-tropic HIV-1_{MN} envelope glycoprotein gp120 as follows: streptavidin-coated fluorescent beads were incubated with biotinylated F(ab')$_2$ fragment rabbit anti–sheep IgG (6 μg/ml; Jackson ImmunoResearch Laboratories) followed by an overnight incubation with sheep anti-gp120 antibody D7324 (Alto Bio Reagents, Ltd.) at 4°C. The beads were washed and incubated with 250 ng/ml purified HIV-1 gp120 (provided by Immunodiagnostics, Inc., through the National Institutes of Health AIDS Research and Reference Reagent Program) overnight at 4°C. The fluorescent beads adhesion assay was performed as described by Geijtenbeek et al. (11). In brief, cells were resuspended in adhesion buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM $CaCl_2$, 2 mM $MgCl_2$, 0.5% BSA) at a final concentration of 5×10^6 cells/ml. 50,000 cells were preincubated with mAb (20 μg/ml) for 10 min at room temperature. Ligand-coated fluorescent beads (20 beads/cell) were added and the suspension was incubated for 30 min at 37°C. Adhesion was determined by measuring the percentage of cells that bound fluorescent beads using flow cytometry on a FACScan™ (Becton Dickinson).

Detection of L-SIGN on Primary Human Liver Sinusoidal Endothelial Cells. Liver tissue was obtained from a patient undergoing liver surgery after having received written consent. Isolation of primary human liver cells was performed as described previously (12). Cells were cultured on collagen type I–coated tissue culture plates in supplemented Williams E Medium (13). The day after isolation, liver cells were incubated with Texas red–labeled OVA (10 μg/ml; Molecular Probes) for 2 h and detached from the matrix by gentle trypsin treatment. Cells were stained with rabbit anti–L-SIGN antiserum followed by goat anti–rabbit Ig FITC (Dianova) and analyzed with a FACScan™ (Becton Dickinson) using CELLQuest™ software. OVA uptake was characteristic of liver sinusoidal endothelial cells (LSECs) only and not Kupffer cells, as verified by the costaining of OVA^+ cells with an endothelial cell–specific marker, acetylated LDL, using confocal microscopy.

HIV-1 Infection Assays. The infection assays were performed as described previously (1, 2). Pseudotyped HIV-1 stocks were generated by calcium phosphate transfections of HEK293T cells with the proviral vector plasmid NL-Luc-E$^-$R$^-$ containing a firefly luciferase reporter gene (14) and expression plasmids for either ADA or JRFL gp160 envelopes. Viral stocks were evaluated by limiting dilution on GHOST CXCR4/CCR5 and 293T-CD4-CCR5 cells. In HIV-1 cell capture assays, DC-SIGN or L-SIGN expressing THP-1 transfectants (250,000 cells) were preincubated with pseudotyped HIV-1 (multiplicity of infection ~0.1 with regard to target cell concentration) in a total volume of 0.5 ml for 3 h to allow cellular adsorption of the virus. After 3 h incubation, cells were washed with 2 vol PBS and the THP-1 transfectants were cocultured with Hut/CCR5 targets (100,000 cells) in the presence of 10 μg/ml polybrene in 1 ml cell culture medium. Cell lysates were obtained after 3 d and analyzed for luciferase activity. In contrast, HIV-1 enhancement assays used suboptimal concentrations of virus (typically <0.05 multiplicity of infection) without a wash step. In brief, *DC-SIGN* or *L-SIGN* transfectants (50,000 cells) were incubated with identical virus concentrations (either pseudotyped HIV-1 or replication-competent M-tropic strain HIV-1_{JR-CSF}), and after 2 h activated T cells (100,000 cells) were added. Cell lysates were obtained after several days and analyzed for either luciferase activity or p24 antigen levels. T cells were activated by culturing them in the presence of 10 U/ml IL-2 and 10 μg/ml PHA for 2 d.

Immunohistochemical Analysis. Staining of the tissue cryosections was performed as described previously (2). 8-μm cryosections of the tissues were fixed in 100% acetone for 10 min, washed with PBS, and incubated with the first antibody (10 μg/ml) for 60 min at 37°C. After washing, the final staining was performed with the ABC-PO/ABC-AP Vectastain kit (Vector Laboratories) according to the manufacturer's protocol. Nuclear staining was performed with hematoxylin.

Results

Genomic Map of DC-SIGN and L-SIGN. A fine map of the *DC-SIGN/L-SIGN* gene locus was determined using information from the human BAC clone CTD-2102F19 sequence, which is now available in GenBank/EMBL/DDBJ (accession no. AC008812; Fig. 1). *DC-SIGN* and *L-SIGN* are positioned in a head-to-head orientation 15.7 kb apart. RH mapping indicated that *DC-SIGN* and *L-SIGN* are located on chromosome 19p13.2-3, near the marker D19S912 (lod score values >11.1) with *DC-SIGN* positioned more telomeric. In agreement with the RH data, the D19S912 marker is found at a distance of ~37 kb centromeric to *L-SIGN* on the BAC sequence.

Soilleux et al. (5) reported a *DC-SIGNR* cDNA clone that contained eight exons with an additional 3′ intron spliced out of the 3′ untranslated region (UTR) compared with the cDNA clone described by Yokoyama-Kobayashi et al. (4). Our RT-PCR experiments on different tissues (liver, pancreas, lung, and placenta) showed that the splice variant described by Soilleux et al. is consistently present but only as a minor transcript, whereas the major transcript consists of seven exons (data not shown). Also, a transcript missing exons 2 and 6, as described by Yokoyama-Kobayashi et al. (4), was extremely rare in our hands.

Northern hybridization data (see below) indicated that *DC-SIGN* mRNA is 3 kb longer than that reported previously (3, 5). We found that this difference is due to the presence of an additional 3-kb UTR in exon 7. Indeed, there was no canonical polyadenylation signal in either of the previously published *DC-SIGN* cDNA sequences, and a search of GenBank/EMBL/DDBJ sequences revealed short polyadenylated ESTs with putative poly(A) signal

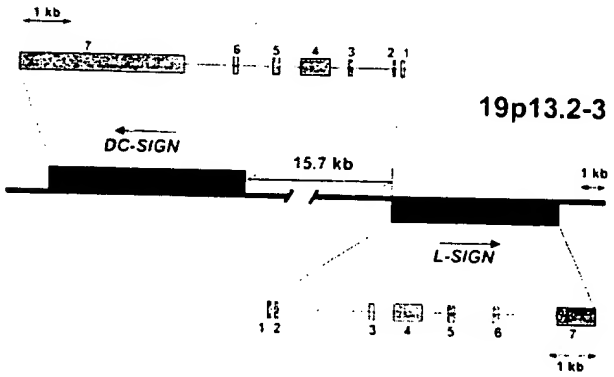


Figure 1. Schematic representation of the *DC-SIGN/L-SIGN* genetic map. Physical distances and gene orientation are based on the sequence provided from BAC clone CTD-2102F19 (GenBank/EMBL/DDBJ accession no. AC008812).

motifs mapping 3 kb downstream of the alleged 3′ end of *DC-SIGN*. RT-PCR experiments indicated that those ESTs correspond to *DC-SIGN* mRNA (data not shown). Based on these findings, we conclude that the full *DC-SIGN* transcript contains the additional 3 kb, resulting in a total of 4.3 kb.

Polymorphism in Exon 4 of L-SIGN. Exon 4 of both *DC-SIGN* and *L-SIGN* contains repeats of 69 bp that encode repeating units of 23 amino acids. These repeats form a neck between the carbohydrate recognition domain and the transmembrane domain of the SIGN molecules. The *L-SIGN* cDNA clone isolated from placental mRNA contained the entire coding region of the gene, but only six full repeats were present in the sequence corresponding to exon 4, in contrast to seven full repeats identified in the cDNA reported by Soilleux et al. (5). This indicated that the repeat region of *L-SIGN* is polymorphic. Analysis of exon 4 in 350 Caucasian individuals showed the presence of seven alleles based on number of repeats (ranging from three to nine), the most common of which was the allele containing seven repeats (Table I). Strikingly, analysis of *DC-SIGN* exon 4 in 150 Caucasians did not reveal any variability.

Northern Blot Analysis of DC-SIGN and L-SIGN. *L-SIGN* mRNA exhibits ~90% similarity to *DC-SIGN* mRNA over the entire coding region, but there is only 53% similarity between exons 2 of the genes. Therefore, exon 2 sequence was used to generate a probe (84 nt) that was *L-SIGN* specific in Northern blot analysis. The probe hybridized to mRNA of ~1.9, 2.6, and 4.2 kb in size in liver and lymph node, and a weak 1.9-kb band was detected in thymus (Fig. 2 A). The 1.9-kb band, which is prominent in lymph node and fetal liver, corresponds to the predicted size of *L-SIGN*. The upper bands (one of which, 2.6 kb, is substantial in adult liver) are likely to be alternative transcripts, but RACE and RT-PCR techniques have not indicated the presence of UTRs varying in length nor alternative splice variants. Therefore, we cannot exclude the possibility that a gene(s) with homology to *L-SIGN* and precisely the same expression pattern is present in humans, but a thorough search for such genes in the sequence databases has been unsuccessful. Finally, a polymorphism in the *L-SIGN* gene (e.g., exon 4 repeat expansion or alteration in the polyadenlation signal motif) could possibly explain the larger transcript size.

Northern blots were reprobed with a 1.2-kb fragment containing the entire coding sequence of *DC-SIGN*, which recognizes both *DC-* and *L-SIGN* mRNA due to their high sequence similarity (Fig. 2 B). Once again, the bands representing *L-SIGN* transcripts were observed in liver, lymph node, and fetal liver. Additionally, a 4.3-kb transcript representing *DC-SIGN* was detected in monocyte-derived DCs and lymph node, and to a lesser extent, in placenta, spleen, thymus, and possibly liver.

L-SIGN mRNA was also detected in placenta and DCs using a more sensitive RT-PCR technique, in agreement with previously reported data (5), but the level of expression in these tissues is too low to be detected by Northern hybridization. The probe which recognizes both *DC-SIGN* and *L-SIGN* transcripts with nearly equal sensitivity clearly indicated differential tissue distribution of the two gene products: *L-SIGN* is primarily transcribed in liver and lymph node, whereas *DC-SIGN* is specifically expressed in DCs and tissues that accommodate DCs (Fig. 2; reference 1). Although both *L-SIGN* and *DC-SIGN* mRNAs are found in lymph node, it is likely that they are expressed by different cell types in this tissue. DCs, which are frequent in lymph node, are the source of *DC-SIGN* mRNA in this tissue (2), but *L-SIGN* mRNA is not detected by Northern blot analysis in DCs, peripheral blood lymphocytes, or spleen (Fig. 2). It is possible that *L-SIGN* expression may be inducible in certain leukocytes during specific stages of activation or, perhaps more likely, endothelial cells of the lymph node may constitutively express this receptor. Char-

Table I. *Polymorphism of the Repeat Region in L-SIGN Exon 4*

No. of repeats	Allele frequency (percent)
3	1 (0.3)
4	25 (3.6)
5	202 (28.9)
6	86 (12.2)
7	377 (53.9)
8	2 (0.3)
9	7 (1.0)

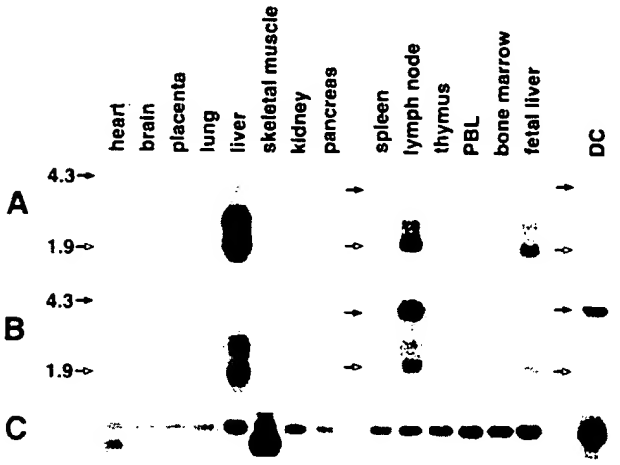


Figure 2. Northern blot analysis of *DC-SIGN* and *L-SIGN*. Positions of the 4.3-kb (black arrows) and 1.9-kb (white arrows) sizes are marked on the left. (A) Hybridization with the *L-SIGN*–specific probe indicating expression of the gene in liver, lymph node, and weakly in thymus. (B) Hybridization with the probe recognizing both genes. 4.3-kb bands represent *DC-SIGN* mRNA. The light upper band (~4.2 kb) evident in liver and lymph node using the *L-SIGN*–specific probe (Fig. 3 A) is distinct from *DC-SIGN* mRNA (4.3 kb) due to the specificity of the probe, intensity patterns, and slight differences in size. (C) Reprobing of the blots with the β-actin cDNA control probe.

acterization of the mechanism involved in the differential tissue expression of these two highly homologous molecules will be of particular interest.

L-SIGN Is Expressed by Human LSECs and Not by DCs. To identify the cells expressing L-SIGN molecules in vivo, we performed immunohistochemical analysis using a pair of anti–DC-SIGN mAbs, one of which, AZN-D3, cross-reacted with L-SIGN, whereas another, AZN-D1, was DC-SIGN specific (Fig. 3 A). As expected from the Northern blot analysis, poor staining of liver tissue was observed using the DC-SIGN–specific mAb AZN-D1 (Fig. 3 B), and the rare cells detected with this antibody are probably DCs residing in liver. In contrast, the mAb AZN-D3 brightly stained cells lining the sinusoids of the liver (Fig. 3 B). mAbs against the endothelial cell–specific marker CD31 gave a similar staining pattern on serial liver sections (data not shown), suggesting that L-SIGN is expressed by LSECs. To support this idea, primary human LSECs were distinguished from the other hepatic cells by uptake of OVA, which is a unique characteristic of LSECs (15), and were tested for expression of L-SIGN directly. Staining of LSECs with polyclonal anti–L-SIGN antibodies indicated that L-SIGN is expressed exclusively by these cells in liver (Fig. 3 C).

Both AZN-D1 and AZN-D3 stained lymph node equally well (data not shown), but without sufficient definition to determine whether cellular staining patterns differed between the two antibodies. However, using L-SIGN–specific polyclonal antibodies, we found that L-SIGN is not expressed by monocyte-derived DCs (Fig. 3 D), which supports conclusions from the Northern blot analysis. Therefore, it is likely, that DC-SIGN and L-SIGN are expressed by different types of cells in the lymph node.

L-SIGN Binds ICAM-3 and HIV-1 gp120. We predicted that L-SIGN and DC-SIGN would bind similar ligands given the nearly identical amino acid sequence of their extracellular domains. Both ICAM-3 and HIV-1_{MN} gp120 have been shown to bind with high affinity to DC-SIGN in a Ca^{2+}-dependent manner (1–3). Using a flow cytometry–based adhesion assay (11), K562 cells transfected

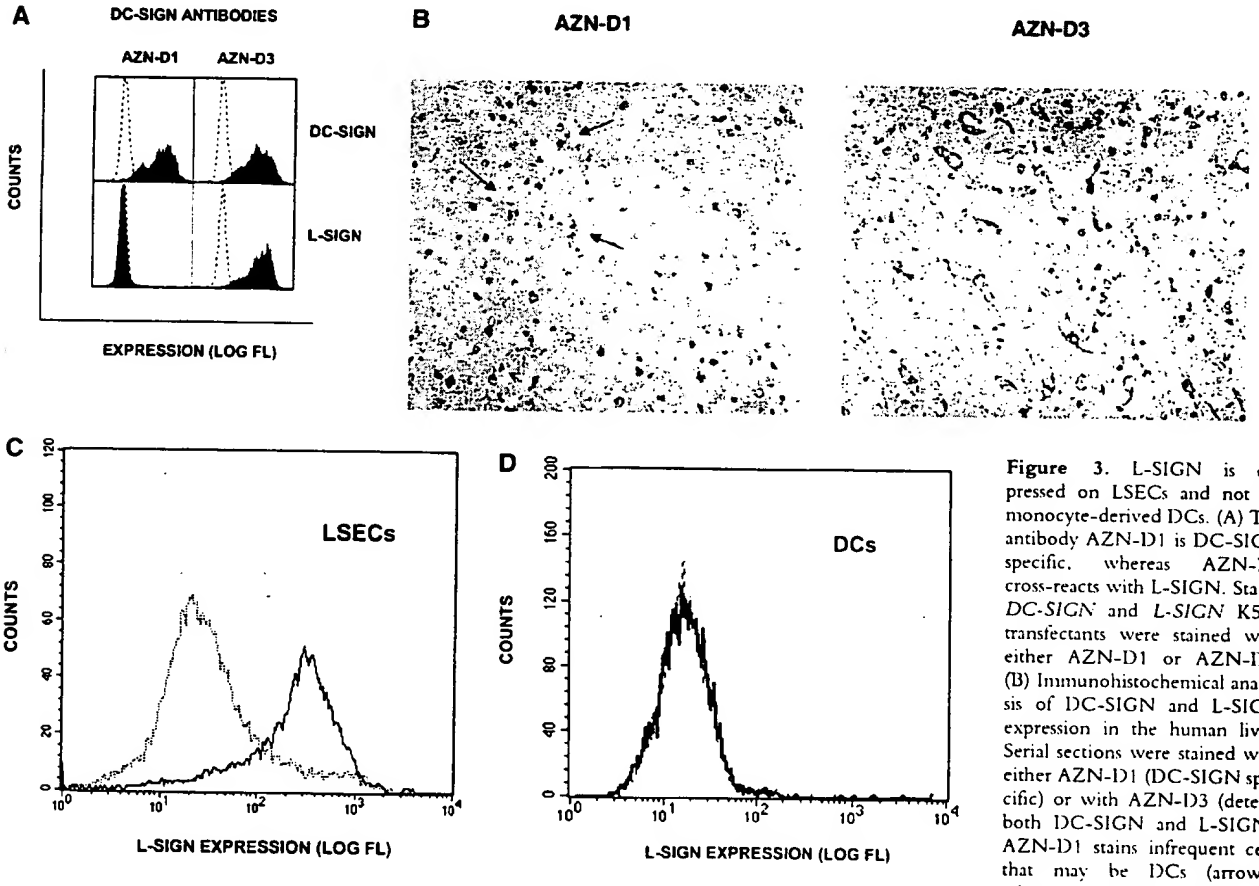


Figure 3. L-SIGN is expressed on LSECs and not on monocyte-derived DCs. (A) The antibody AZN-D1 is DC-SIGN specific, whereas AZN-D3 cross-reacts with L-SIGN. Stable *DC-SIGN* and *L-SIGN* K562 transfectants were stained with either AZN-D1 or AZN-D3. (B) Immunohistochemical analysis of DC-SIGN and L-SIGN expression in the human liver. Serial sections were stained with either AZN-D1 (DC-SIGN specific) or with AZN-D3 (detects both DC-SIGN and L-SIGN). AZN-D1 stains infrequent cells that may be DCs (arrows), whereas AZN-D3 stains cells lining sinusoids. (C) Expression of L-SIGN in liver is restricted to LSECs. 1 d after isolation, primary human liver cells were incubated with fluorochrome-labeled OVA. L-SIGN expression was determined by indirect immunofluorescence using an L-SIGN–specific polyclonal antibody. Cells that have taken up OVA (LSECs) and those that did not take up OVA (hepatocytes and other resident hepatic cells) are represented by solid and broken lines, respectively, by gating on the respective cell populations. 2×10^5 cells were analyzed. (D) L-SIGN is not expressed by monocyte-derived DCs. Immature DCs, cultured from monocytes in the presence of GM-CSF and IL-4, do not stain with anti–L-SIGN polyclonal antibody, as determined by FACScan™ analysis. Solid line indicates staining with anti–L-SIGN polyclonal serum, whereas broken line (hidden under solid lane) represents staining with rabbit preimmune serum.

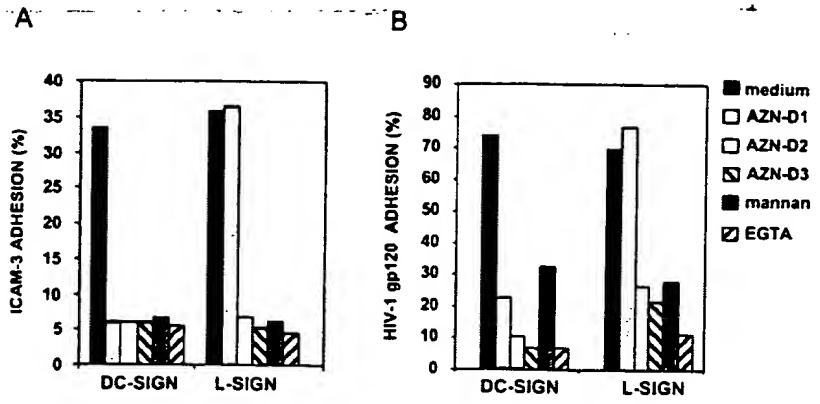


Figure 4. L-SIGN binds ICAM-3 (A) and HIV-1 gp120 (B). Adhesion of ICAM-3 and gp120 to the K562–L-SIGN and K562–DC-SIGN cells was measured with the fluorescent bead adhesion assay (reference 11). The y-axis represents the percentage of cells binding ligand-coated fluorescent beads. The L-SIGN cross-reacting mAbs AZN-D2 (20 μg/ml) and AZN-D3 (20 μg/ml) inhibit the adhesion of ICAM-3 and gp120 to L-SIGN, in contrast to the DC-SIGN–specific mAb AZN-D1 (20 μg/ml). Adhesion of both ICAM-3 and gp120 to the K562 transfectants is also inhibited by either 20 μg/ml mannan or 5 mM EGTA. Adhesion of both ligands to mock transfectants was <5%. One representative experiment out of three is shown (SD < 5%).

with *L-SIGN* were shown to bind ICAM-3 with high affinity (Fig. 4 A). The L-SIGN–mediated binding was inhibited by the DC-SIGN/L-SIGN–specific mAbs AZN-D2 and AZN-D3, mannan, or EGTA, but not by the DC-SIGN–specific mAb AZN-D1, demonstrating that L-SIGN functions as a mannose-binding C-type lectin with a high affinity for ICAM-3. As predicted by the high homology to DC-SIGN, L-SIGN was also able to bind to HIV-1_{MN} gp120 in a manner similar to that observed for DC-SIGN (1) (Fig. 4 B). Mock transfected cells did not bind either ICAM-3 or HIV-1_{MN} gp120 (data not shown).

L-SIGN Enhances HIV-1 Infection. High affinity binding of L-SIGN to HIV-1 gp120 raised the possibility that, like DC-SIGN, L-SIGN might bind infectious HIV-1 and enhance infection of target cells in trans. To test the role of L-SIGN as a transreceptor in HIV-1 infection, THP-1 cells expressing either DC-SIGN or L-SIGN were pulsed with single round infectious HIV-luciferase pseudotyped with M-tropic HIV-1_{JRFL} envelope glycoprotein, washed to remove unbound virus, and incubated with target cells permissive for HIV-1 infection. Infection was evaluated after 3 d. Both the *L-SIGN*– and *DC-SIGN*–transfected THP-1 cells captured infectious HIV-1 and transmitted the virus to target cells, while mock transfected THP-1 cells did not (Fig. 5 A).

Next we investigated whether L-SIGN would be able to capture a limiting concentration of HIV-1 and efficiently present the virus to the permissive cells promoting infection. HEK293T cells expressing DC-SIGN or L-SIGN, or mock

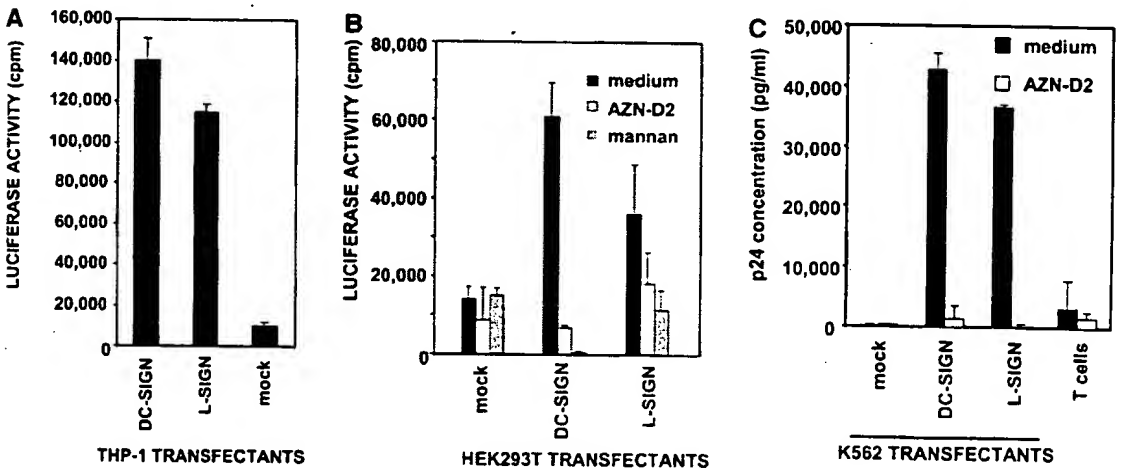


Figure 5. L-SIGN captures and enhances infection of T cells with HIV-1 in trans. (A) L-SIGN captures HIV-1 and transmits it to target cells. Stable DC-SIGN– or L-SIGN–expressing THP-1 transfectants were preincubated with HIV-luc/JRFL pseudovirions to allow capture of the virus. Cells were washed and THP-1 transfectants were cocultured with Hut/CCR5 target cells. Cell lysates were obtained after 3 d and analyzed for luciferase activity. For each of the coculture conditions employed, mock infected controls were uniformly <100 cps in activity. Each data set represents the mean of four separate wells of infected cells. One representative experiment out of two is shown. (B) L-SIGN enhances infection of T cells by pseudotyped HIV-1. HEK293T cells were transiently transfected with cDNA encoding DC-SIGN, L-SIGN, or empty vector. Control cells were preincubated with 20 μg/ml AZN-D2 or 20 μg/ml mannan. Low amounts of pseudotyped HIV-1_{ADA} were added together with activated T cells as described previously (reference 1). Infectivity was determined after 2 d by measuring luciferase activity. One representative experiment of two performed is shown. Each experiment was done in triplicate wells. (C) L-SIGN enhances infection of T cells by replication competent HIV-1. Stable K562 transfectants of both L-SIGN and DC-SIGN were incubated with low virus concentrations of replication-competent M-tropic strain HIV-$1_{JR\text{-}CSF}$ ($TCID_{50}$ 100/ml). To determine the specificity, cells were preincubated with AZN-D2 (20 μg/ml). After 2 h, activated T cells were added as described previously (reference 2). Culture supernatants were collected at day 14 after K562-T cell coculture and HIV-1 production was measured using ELISA to determine p24 antigen levels. In control experiments, the same amount of virus was added directly to T cells. One representative experiment out of three is shown. Each data set represents the mean of three separate wells of infected cells.

transfected cells were incubated with low titers of HIV-luciferase pseudotyped with HIV-1_{ADA} envelope glycoprotein. The unwashed cells were then cocultured with activated T cells. Minimal infection of target cells was observed from mock transfected HEK293T cells pulsed with HIV-1 (Fig. 5 B). However, HEK293T cells transfected with L-SIGN enhanced HIV-1 infection of T cells in trans, similar to DC-SIGN (Fig. 5 B). The DC-SIGN–mediated enhancement was inhibited with the cross-reactive AZN-D2 antibody, while partial inhibition was observed for L-SIGN, possibly because of some difference in the reactivity of this antibody to the two SIGN molecules that was evident under the conditions employed in this experiment. Mannan efficiently inhibited enhancement by both SIGN molecules.

Similar experiments to evaluate the ability of L-SIGN to enhance HIV-1 infection of T cells were performed using replication-competent virus. K562 cells transfected with *L-SIGN*, *DC-SIGN*, and empty vector were incubated with the M-tropic HIV-1_{JR-CSF} strain at low virus concentrations for 2 h and subsequently cocultured with activated T cells (Fig. 5 C). No viral replication was observed using mock transfected K562 cells, while *L-SIGN* transfectants transmitted HIV-1 to target cells, resulting in viral replication with nearly the same efficiency as *DC-SIGN* transfectants. Almost complete inhibition of HIV-1 replication with the DC-SIGN/L-SIGN–specific antibody AZN-D2 indicated the specificity of these receptors to enhance HIV-1 infection. Thus, non-DC lineage cells expressing L-SIGN within liver and possibly in lymph node may also have the ability to capture and transmit HIV-1 to lymphocytes.

Discussion

The homologous human C-type lectins DC-SIGN and L-SIGN appear to be the products of a recent gene duplication. The corresponding proteins share the same domain organization and overlapping, if not completely identical, ligand specificity. The most diverse region of these molecules occurs in their cytoplasmic tails (5). It has been suggested that DC-SIGN–associated HIV-1 may be internalized, protecting it from degradation or inactivation (1). If so, the sequence variation in the cytoplasmic region of L-SIGN relative to DC-SIGN could affect the level of receptor internalization and viral uptake, perhaps explaining the consistent differences in efficiency of HIV-1 infection enhancement observed in our experiments between *DC-SIGN* and *L-SIGN* transfectants (Fig. 5).

Another obvious difference in *SIGN* genes is the repeat polymorphism in exon 4 of *L-SIGN*, which is conserved in *DC-SIGN* (Table I). The neck domain of L-SIGN may contain from three to nine repeats, while DC-SIGN always consists of seven repeats among the Caucasians tested. It is not clear whether the differences in exon 4 diversity of these genes is because of some distinction in the physical feature(s) of the genes or to selective processes acting on the genes differentially. The neck domain may be involved in oligomerization of the receptors (5) and variable numbers of repeats could potentially affect functional characteristics of the L-SIGN molecule, particularly in heterozygotes where heterooligomers might be present. However, our preliminary data indicated no difference between L-SIGN molecules containing six or seven repeats in ligand binding or in HIV-1 capture and enhancement experiments.

Although the *SIGN* genes have maintained sequence and functional similarity over their evolutionary history, regulatory elements determining their tissue distribution have evolved along unique paths. Northern blot analysis of mRNA expression clearly indicated expression of *DC-SIGN* in monocyte-derived DCs and in tissues where DCs reside, whereas expression of *L-SIGN* in DCs was undetectable (Fig. 2). Further, L-SIGN was not detected on monocyte-derived DCs using antibodies specific to L-SIGN (Fig. 3 C). Thus, it is most likely that unique cell types in the lymph node express one but not both SIGN molecules: L-SIGN could be expressed by endothelial cells, as it is in liver, whereas DC-SIGN is expressed by DCs in the T cell area of lymph node (2).

Liver sinusoids are specialized capillary vessels characterized by the presence of resident macrophages adhering to the endothelial lining. The LSEC–leukocyte interactions, which require expression of adhesion molecules on the cell surfaces, appear to constitute a central mechanism of peripheral immune surveillance in the liver (15). The mannose receptor as well as other costimulatory receptors such as MHC class II, CD80, and CD86, are known to be expressed on LSECs and to mediate the clearance of many potentially antigenic proteins from the circulation in a manner similar to DCs in lymphoid organs (15). L-SIGN may fit in this category of receptors on LSECs, as its tissue location and ligand-binding properties strongly implicate a physiologic role for this receptor in antigen clearance, as well as in LSEC–leukocyte adhesion. The high expression of ICAM-3 on apoptotic cells (16) may provide the means by which these cells are trapped by L-SIGN–expressing cells in the liver and subsequently cleared.

The mannose glycans present on gp120 appear to mediate HIV-1 adhesion to the SIGN molecules, although the contribution of the gp120 polypeptide backbone is not excluded (1). Several in vitro studies have shown that highly glycosylated HIV-1 gp120 is a strong ligand for a variety of mannose-binding lectins (17–22). Although the carbohydrate structures on the HIV envelope could be nonspecifically recognized by host lectins, the physiological consequences of such recognition will be specified by the functions of the binding molecule. The SIGN molecules are the first membrane-associated lectins identified to date that enhance HIV-1 infection. Interestingly, the expression of L-SIGN in liver sinusoids suggests that LSECs, which are in continual contact with passing leukocytes, can capture HIV-1 from the blood and promote transinfection of T cells. Moreover, prior studies have indicated that LSECs themselves may be susceptible to HIV-1 infection (23, 24). Thus, it is possible that L-SIGN promotes infection of these cells, thereby establishing a reservoir for production of a new virus to pass on to T lymphocytes trafficking through the liver sinusoid.

Additional functional studies are necessary for understanding the normal physiologic role of L-SIGN and its possible role in HIV-1 pathogenesis. Its ability to enhance transinfection of T cells suggests that L-SIGN may contribute to HIV-1 susceptibility. Alternatively, if a physiologic function of L-SIGN involves antigen clearance, this receptor could play a protective role in clearance of the virus from the circulation. A clearer understanding of this receptor may provide insight into its potential use in novel therapy against HIV-1.

We thank Y. Liu, P. Karacki, and M. Chalaby for their technical contributions and the laboratory of J. Goudsmit for providing support. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-56000, and the Dutch Aids Foundation contract 5008.

Submitted: 22 December 2000
Revised: 31 January 2001
Accepted: 5 February 2001

References

1. Geijtenbeek, T.B., D.S. Kwon, R. Torensma, S.J. van Vliet, G.C. van Duijnhoven, J. Middel, I.L. Cornelissen, H.S. Nottet, V.N. KewalRamani, D.R. Littman, et al. 2000. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell.* 100:587–597.
2. Geijtenbeek, T.B., R. Torensma, S.J. van Vliet, G.C. van Duijnhoven, G.J. Adema, Y. van Kooyk, and C.G. Figdor. 2000. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell.* 100:575–585.
3. Curtis, B.M., S. Scharnowske, and A.J. Watson. 1992. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc. Natl. Acad. Sci. USA.* 89:8356–8360.
4. Yokoyama-Kobayashi, M., T. Yamaguchi, S. Sekine, and S. Kato. 1999. Selection of cDNAs encoding putative type II membrane proteins on the cell surface from a human full-length cDNA bank. *Gene.* 228:161–167.
5. Soilleux, E.J., R. Barten, and J. Trowsdale. 2000. DC-SIGN, a related gene DC-SIGNR, and CD23 form a cluster on 19p13. *J. Immunol.* 165:2937–2942.
6. Chomczynski, P. 1992. One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Anal. Biochem.* 201:134–139.
7. Sallusto, F., and A. Lanzavecchia. 1994. Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor α. *J. Exp. Med.* 179:1109–1118.
8. Romani, N., S. Gruner, D. Brang, E. Kampgen, A. Lenz, B. Trockenbacher, G. Konwalinka, P.O. Fritsch, R.M. Steinman, and G. Schuler. 1994. Proliferating dendritic cell progenitors in human blood. *J. Exp. Med.* 180:83–93.
9. Lub, M., S.J. van Vliet, S.P. Oomen, R.A. Pieters, M. Robinson, C.G. Figdor, and Y. van Kooyk. 1997. Cytoplasmic tails of β1, β2, and β7 integrins differentially regulate LFA-1 function in K562 cells. *Mol. Biol. Cell.* 8:719–728.
10. Cecilia, D., V.N. KewalRamani, J. O'Leary, B. Volsky, P. Nyambi, S. Burda, S. Xu, D.R. Littman, and S. Zolla-Pazner. 1998. Neutralization profiles of primary human immunodeficiency virus type 1 isolates in the context of coreceptor usage. *J. Virol.* 72:6988–6996.
11. Geijtenbeek, T.B., Y. van Kooyk, S.J. van Vliet, M.H. Renes, R.A. Raymakers, and C.G. Figdor. 1999. High frequency of adhesion defects in B-lineage acute lymphoblastic leukemia. *Blood.* 94:754–764.
12. Hegenbarth, S., R. Gerolami, U. Protzer, P.L. Tran, C. Brechot, G. Gerken, and P.A. Knolle. 2000. Liver sinusoidal endothelial cells are not permissive for adenovirus type 5. *Hum. Gene Ther.* 11:481–486.
13. Hild, M., O. Weber, and H. Schaller. 1998. Glucagon treatment interferes with an early step of duck hepatitis B virus infection. *J. Virol.* 72:2600–2606.
14. Connor, R.I., B.K. Chen, S. Choe, and N.R. Landau. 1995. Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. *Virology.* 206:935–944.
15. Knolle, P.A., and G. Gerken. 2000. Local control of the immune response in the liver. *Immunol. Rev.* 174:21–34.
16. Moffatt, O.D., A. Devitt, E.D. Bell, D.L. Simmons, and C.D. Gregory. 1999. Macrophage recognition of ICAM-3 on apoptotic leukocytes. *J. Immunol.* 162:6800–6810.
17. Ezekowitz, R.A., M. Kuhlman, J.E. Groopman, and R.A. Byrn. 1989. A human serum mannose-binding protein inhibits in vitro infection by the human immunodeficiency virus. *J. Exp. Med.* 169:185–196.
18. Gattegno, L., A. Ramdani, T. Jouault, L. Saffar, and J.C. Gluckman. 1992. Lectin-carbohydrate interactions and infectivity of human immunodeficiency virus type 1 (HIV-1). *AIDS Res. Hum. Retroviruses.* 8:27–37.
19. Hammar, L., I. Hirsch, A. Machado, J. De Mareuil, J. Baillon, and J.C. Chermann. 1994. Lectin effects on HIV-1 infectivity. *Ann. NY Acad. Sci.* 724:166–169.
20. Saifuddin, M., M.L. Hart, H. Gewurz, Y. Zhang, and G.T. Spear. 2000. Interaction of mannose-binding lectin with primary isolates of human immunodeficiency virus type 1. *J. Gen. Virol.* 81:949–955.
21. Ushijima, H., H.C. Schroder, S. Poznanovic, M.J. Gasic, E. Matthes, and W.E. Muller. 1992. Inhibition of human immunodeficiency virus-1 infection by human conglutinin-like protein: in vitro studies. *Jpn. J. Cancer Res.* 83:458–464.
22. Milone, M.C., and P. Fitzgerald-Bocarsly. 1998. The mannose receptor mediates induction of IFN-α in peripheral blood dendritic cells by enveloped RNA and DNA viruses. *J. Immunol.* 161:2391–2399.
23. Steffan, A.M., M.E. Lafon, J.L. Gendrault, C. Schweitzer, C. Royer, D. Jaeck, J.P. Arnaud, M.P. Schmitt, A.M. Aubertin, and A. Kirn. 1992. Primary cultures of endothelial cells from the human liver sinusoid are permissive for human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA.* 89:1582–1586.
24. Housset, C., E. Lamas, V. Courgnaud, O. Boucher, P.M. Girard, C. Marche, and C. Brechot. 1993. Presence of HIV-1 in human parenchymal and non-parenchymal liver cells in vivo. *J. Hepatol.* 19:252–258.

JOURNAL OF VIROLOGY, July 2002, p. 6841–6844
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.13.6841–6844.2002

Vol. 76, No. 13

C-Type Lectins DC-SIGN and L-SIGN Mediate Cellular Entry by Ebola Virus in *cis* and in *trans*

Carmen P. Alvarez,[1] Fátima Lasala,[1] Jaime Carrillo,[1]† Oscar Muñiz,[2] Angel L. Corbí,[2] and Rafael Delgado[1]*

Laboratory of Molecular Microbiology, Hospital 12 de Octubre,[1] and Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas,[2] Madrid, Spain

Received 4 January 2002/Accepted 29 March 2002

Ebola virus is a highly lethal pathogen responsible for several outbreaks of hemorrhagic fever. Here we show that the primate lentiviral binding C-type lectins DC-SIGN and L-SIGN act as cofactors for cellular entry by Ebola virus. Furthermore, DC-SIGN on the surface of dendritic cells is able to function as a *trans* receptor, binding Ebola virus-pseudotyped lentiviral particles and transmitting infection to susceptible cells. Our data underscore a role for DC-SIGN and L-SIGN in the infective process and pathogenicity of Ebola virus infection.

Ebola virus is responsible for several major outbreaks of hemorrhagic fever, the exceedingly high mortality of which has raised great public concern. Ebola virus research has been hampered by the strict biosafety containment procedures required for handling the infectious agent. However, the structural similarity of Ebola virus glycoprotein (GP) to retroviral envelopes (6) has recently allowed the generation of pseudotyped recombinant retroviral particles that have been used to explore important aspects of the Ebola virus biology (16, 18). Ebola virus cell entry is presumably mediated by the interaction of a cellular receptor with the GP1 subunit of the viral envelope (12). A cofactor for cellular entry of Ebola virus and Marburg filoviruses in certain cell types has been recently identified as the folate receptor α (FRα) (3). This molecule is a glycophosphatidylinositol-linked protein highly conserved in mammalian species and expressed in epithelial and parenchymal cells of a number of organs, but not abundantly in liver or endothelial cells (15).

DC-SIGN (dendritic cell [DC]-specific ICAM-3 grabbing non-integrin, CD209) is a type II membrane protein with a C-type lectin extracellular domain, the expression of which is restricted to immature DC. DC-SIGN appears to play a key role in the initial stages of immune response and in the migratory behavior of DC, because it mediates DC interactions with T lymphocytes and endothelial cells through recognition of ICAM-3 (9) and ICAM-2 (7). DC-SIGN, originally cloned as a human immunodeficiency virus (HIV) gp120-binding protein (5), does not act as a receptor for cellular entry of HIV; instead, it confers to DC the ability to facilitate infection in *trans* of susceptible cells (8). Recently, DC-SIGN and the newly described DC-SIGN homologue L-SIGN have been shown to bind most lentiviruses of primates: HIV-1 (both R5 and X4 strains), HIV-2, and simian immunodeficiency virus (SIV) (13). Unlike DC-SIGN, L-SIGN is not expressed by DC, but is expressed on the surface of endothelial cells in the liver, lymph node sinuses, and placental villi (2). The affinity of these membrane receptors for retroviral GP and their tissue distribution pattern prompted us to study their potential role as binding and entry cofactors for Ebola virus.

To investigate the participation of DC-SIGN in Ebola virus infection, we have utilized lentiviral particles pseudotyped with Ebola virus GP according to a transient transfection protocol previously described (17). The lentiviral vector pNL4-3.Luc.R^-E^-[10] was used for production of vesicular stomatitis virus G (VSV-G) and Ebola virus Zaire and Reston GP pseudotypes. Expression plasmids for the GP of the Zaire and Reston strains of Ebola virus were kindly provided by A. Sanchez, Centers for Disease Control and Prevention (18). Supernatants were obtained 48 h after transfection, filtered (0.45-μm pore size), and stored frozen at −80°C. Infectious titers were estimated by serial dilution on HeLa cells and were typically in the range of 10^7 infectious units/ml for VSV-G and 10^5 infectious units/ml for Ebola virus GP pseudotypes. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, National Institute for Allergy and Infectious Diseases: DC-SIGN and L-SIGN monoclonal antibody DC28 (0.8 mg/ml as ascitic fluid) from F. Baribaud, S. Pöhlmann, J. A. Hoxie, and R. W. Doms (1); pcDNA3-L-SIGN6 from Mary Carrington; and pNL4-3.Luc.R^-E^- from Nathaniel Landau (10).

To investigate the role of DC-SIGN in Ebola virus binding and cellular entry, we first used a stable transfectant of DC-SIGN in the erythroleukemic K562 cell line (14). K562 cells were incubated overnight in 24-well plates with supernatants containing Ebola virus GP-pseudotyped lentivirus at a multiplicity of infection (MOI) of 0.1. Infectivity was measured 48 h after infection by luciferase assay with reagents from Promega (Madison, Wis.) in a Berthold Sirius luminometer (Berthold, Munich, Germany) with a dynamic range from 10^2 to 10^7 relative light units (RLU). Infectivity of the parental K562 cells with an Ebola virus GP-pseudotyped lentiviral construction was detectable, although relatively low. In contrast, infectivity of the DC-SIGN transfectant cell line was 1 order of magnitude higher, and it was significantly reduced in the presence of

* Corresponding author. Mailing address: Laboratory of Molecular Microbiology, Department of Microbiology, Hospital 12 de Octubre, 28041 Madrid, Spain. Phone: 34 91 3908428. Fax: 34 91 5652765. E-mail: rdelgado@hdoc.insalud.es.

† Present address: Instituto de Investigaciones Biomédicas, Universidad Autónoma de Madrid, Madrid, Spain.

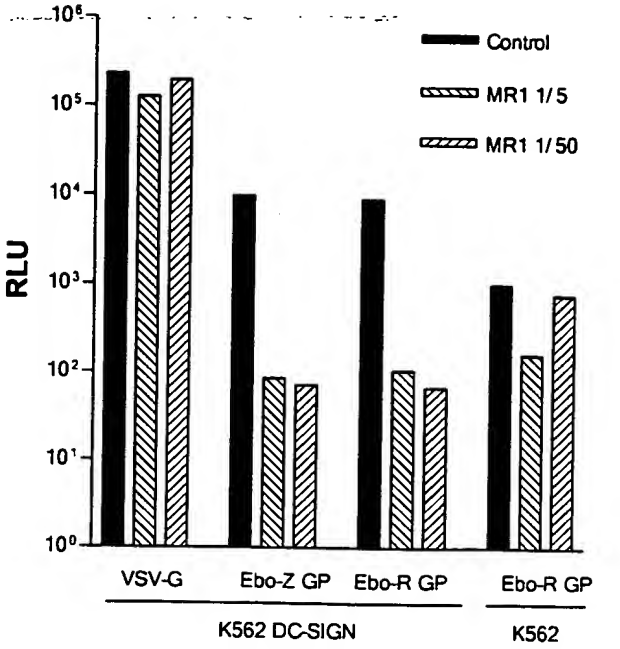


FIG. 1. DC-SIGN-mediated infection of K562-DC-SIGN cells. K562 and K562-DC-SIGN cells were infected with VSV-G-, Ebola virus Zaire (Ebo-Z)-, or Ebola virus Reston (Ebo-R) GP-pseudotyped lentivirus in the absence (control) or presence of the DC-SIGN-specific monoclonal antibody MR-1. Infectivity was measured as luciferase activity 48 h postinfection. One representative experiment out of three is shown.

the DC-SIGN-specific monoclonal antibody MR-1, thus suggesting that Ebola virus might interact with DC-SIGN and facilitate viral entry into K562-DC-SIGN-transfected cells (Fig. 1). MR-1 was used as tissue culture supernatant (10 μg/ml) and showed no reactivity with HeLa and K562 cells, as well as a panel of myeloid and lymphoid cell lines (14).

To further characterize the role of DC-SIGN and its close homologue L-SIGN in Ebola virus cell entry, we expressed, by using retroviral vectors, DC-SIGN and L-SIGN in the Jurkat cell line, since these cells are nonpermissive for Ebola virus infection and are considered receptor deficient (16). Recombinant retroviruses were produced as described previously (17) by cotransfection of the plasmids pNGVL-MLV-gag-pol and pCMV-VSV-G and the retroviral vector pLZRs-DC-SIGN-gfp—constructed by subcloning the DC-SIGN coding sequence obtained from placental RNA by reverse transcription-PCR with primers AAA AGG ATC CGC CGC CAC CAT GAG TGA CTC CAA GGA ACC (forward) and AAA AGA ATT CCT ACG CAG GAG GGG GGT TT (reverse), into the bicistronic retroviral vector pLZRs-M10-gfp (17), digested with *Bam*HI and *Eco*RI—or pLZRs-L-SIGN-gfp constructed in a similar way with the L-SIGN coding sequence obtained from pcDNA3-L-SIGN6. Plasmids pNGVL-MLV-gag-pol, pLZRs-RevM10-gfp, and pCMV-VSV-G were generously provided by G. Nabel, University of Michigan (17). Jurkat cells were transduced with VSV-G-pseudotyped DC-SIGN- or L-SIGN-expressing retroviral vectors by spinoculation for 2 h at 1,500 × *g* at an MOI of 10. After 48 h, cells were analyzed by fluorescence-activated cell sorting for green fluorescent protein (GFP) and lectin expression (range of positive cells, 10 to 30%) and challenged in 24-well plates with Ebola virus GP pseudotypes or controls: 250,000 cells were resuspended in 250 μl of complete medium (RPMI, 10% fetal bovine serum [FBS]) and incubated overnight with 250 μl of supernatant from transfections. Cells were assayed for luciferase expression 48 h postinfection. For inhibition experiments, cells were preincubated for 10 min at room temperature with the carbohydrate-interaction inhibitor mannan (25 μg/ml; Sigma, St. Louis, Mo.) or lectin-specific antibodies. Jurkat cells expressing DC-SIGN or L-SIGN were clearly infected by Ebola virus Zaire and Reston GP-pseudotyped lentiviral vectors, indicating that expression of either of these two lectins in Jurkat cells is sufficient to confer permissivity (Fig. 2A). The DC-SIGN and L-SIGN dependency of the Jurkat cell infection was confirmed

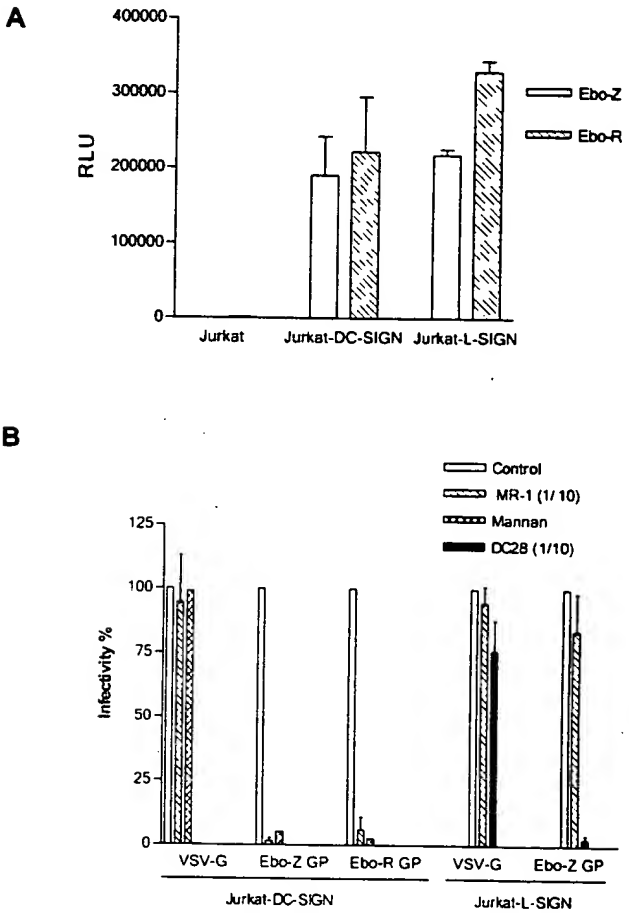


FIG. 2. (A) Jurkat cells expressing DC-SIGN and L-SIGN are permissive for Ebola virus infection. Control Jurkat cells or Jurkat cells expressing DC-SIGN or L-SIGN by transduction with a retroviral vector were infected with Ebola virus Zaire (Ebo-Z) or Reston (Ebo-R) GP-pseudotyped lentivirus (mean ± standard error, $n = 3$). (B) Specificity of DC-SIGN- and L-SIGN-mediated infectivity of Jurkat cells. DC-SIGN- and L-SIGN-mediated infectivity of Jurkat cells transduced with the retroviral vectors mentioned above was assessed by preincubation with mannan and specific monoclonal antibodies: MR-1 is DC-SIGN specific, and DC28 exhibits specificity for both DC-SIGN and L-SIGN. Results are shown as the percentage of luciferase activity compared to that of the untreated cells (mean ± standard error, $n = 3$). Mannan was tested once on Jurkat-DC-SIGN. DC28 was used only for Jurkat L-SIGN.

by the clear reduction of infectivity in the presence of mannan and anti-DC-SIGN and anti-L-SIGN antibodies, whereas a VSV-G-pseudotyped control was unaffected (Fig. 2B). Our results clearly indicate that DC-SIGN and L-SIGN are implicated in Ebola virus GP-mediated cell infection; however, the contribution and the specific molecular interactions of DC-SIGN and L-SIGN in Ebola virus cell entry remain to be defined. In this respect, and since many cells known to be susceptible to Ebola virus do not express these lectins, our results, like those recently reported for HIV and SIV (11), support the hypothesis that DC-SIGN and L-SIGN bind and concentrate Ebola virus to the cell membrane, thus facilitating the interaction in *cis* with cofactors required for cell entry, the low density of which may be limiting for infection of certain cell types.

Finally, the role of DC-SIGN–Ebola virus GP interaction on DC was explored by using monocyte-derived DC (MDDC). MDDC were obtained from blood monocytes according to a standardized protocol (14). Cells were cultured for 5 to 7 days in the presence of granulocyte-macrophage colony-stimulating factor and interleukin-4 to obtain a population of immature MDDC. DC were infected with the lentiviruses pseudotyped with VSV-G and Ebola virus GP (MOI of 10 and 0.1, respectively). Forty-eight to 72 h postinfection, cells were assayed for luciferase expression as described before. Infection of MDDC, although at a low level, was demonstrated by using a VSV-G-pseudotyped control. However, under the conditions used in our experiments and in spite of the high DC-SIGN expression of MDDC, we were unable to readily detect luciferase expression upon infection with Ebola virus GP-pseudotyped lentiviral vectors (data not shown). In this respect, and taking into account the evidence of Ebola virus infection of DC in vitro and in vivo (4), it is possible that limitations of the lentivirus pseudotyping approach, such as low titers or the requirement of additional viral products for entry into DC, might account for this negative result. We next tested whether DC-SIGN on the surface of DC could bind Ebola virus GP-pseudotyped viral particles and facilitate subsequent infection of susceptible cells (Fig. 3). DC were preincubated (150,000 cells in 100 μl) for 20 min at room temperature in the presence or absence of the DC-SIGN-specific antibody MR-1. Supernatants (300 μl) containing Ebola virus GP- or VSV-G-pseudotyped lentiviral particles were then added, and cells were maintained in rotation at room temperature for 2 h. Cells were washed four times in phosphate-buffered saline (PBS)–2% FBS, resuspended in 300 μl of fresh medium, and added to HeLa cells plated in 24-well plates. The same amount of supernatant maintained at room temperature without DC was used as control of infectivity. After 48 h of cocultivation, wells were washed twice with PBS, and HeLa cells were assayed for luciferase activity as described above. The infectivity achieved by cocultivation of HeLa and MDDC, incubated with a high-titer VSV-G-pseudotyped lentiviral supernatant and extensively washed, was more than 2 orders of magnitude lower than that of the initial non-cell-incubated supernatant. The remaining infectivity was unaffected by preincubation of MDDC with a DC-SIGN-specific antibody suggesting that it was most likely due to unspecific binding. In contrast, MDDC incubated with infectious supernatants of Ebola virus GP-pseudotyped viruses retained a higher proportion of the infectivity of the supernatant after extensive washing. This effect was significantly reduced by preincubating MDDC with a DC-SIGN-specific antibody, indicating that MDDC are capable, through DC-SIGN interactions, of binding Ebola virus GP-pseudotyped viruses, maintaining infectivity, and achieving efficient infection in *trans* of susceptible cells in a way similar to that described for lentiviruses (8).

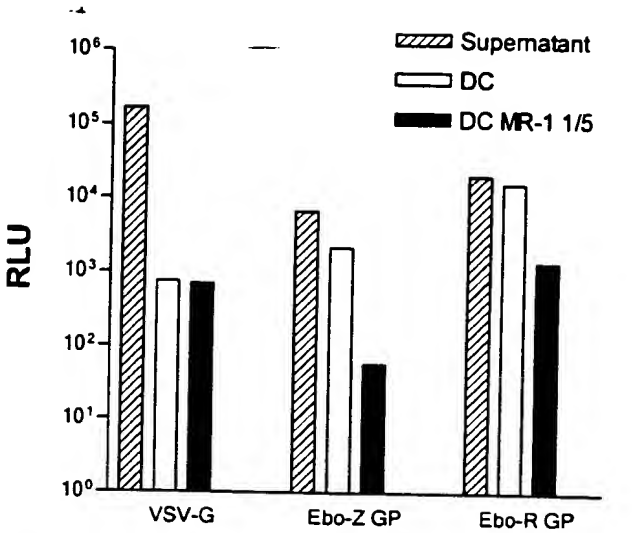


FIG. 3. MDDC bind Ebola virus GP-pseudotyped particles and transmit infectivity to susceptible cells. MDDC were incubated with infectious supernatants containing VSV-G-, Ebola virus Zaire (Ebo-Z)-, or Ebola virus Reston (Ebo-R) GP-pseudotyped lentiviruses after a brief preincubation in the absence or presence of MR-1 DC-SIGN-specific monoclonal antibody. Cells were extensively washed thereafter and plated onto HeLa cells. The same amount of infectious supernatant (Sup) without incubation with DC was directly added to the HeLa cells as a control of the original infectivity. Cells were assayed for luciferase 48 h after infection. The experiment was performed with cells from two independent donors, and a representative result is shown.

We have found that expression of DC-SIGN and its homologue L-SIGN enhances infectivity of Ebola virus-susceptible cells and is sufficient to confer permissivity for Ebola virus GP-mediated infection to a nonsusceptible cell line. Also, DC-SIGN on the surface of DC appears to act as a *trans* receptor capable of binding Ebola virus GP-pseudotyped viruses and efficiently transmitting the infection to susceptible cells. DC-SIGN and L-SIGN appear to be universal binding factors for primate lentiviruses. Our data indicate that these molecules have extended participation in other viral infections. The role of these C-type lectins in Ebola virus primary infection and dissemination deserves further investigation.

We thank Keith Martin for review of the manuscript.

This work has been partially supported by grants CAM 08.3/0026.1/2000 and FIS 01/0063-01 to A.L.C. and FIS 99/0514 and 01/1430 to R.D.

REFERENCES

1. Baribaud, F., S. Pohlmann, T. Sparwasser, M. T. Y. Kimata, Y.-K. Choi, B. S. Haggarty, N. Ahmad, T. Macfarlan, T. G. Edwards, G. J. Leslie, J. Arnason, T. A. Reinhart, J. T. Kimata, D. R. Littman, J. A. Hoxie, and R. W. Doms. 2001. Functional and antigenic characterization of human, rhesus macaque, pigtailed macaque, and murine DC-SIGN. J. Virol. 75:10281–10289.
2. Bashirova, A. A., T. B. H. Geijtenbeek, G. C. F. van Duijnhoven, S. J. van Vliet, J. B. G. Eilering, M. P. Martin, L. Wu, T. D. Martin, N. Viebig, P. A. Knolle, V. N. KewalRamani, Y. van Kooyk, and M. Carrington. 2001. A

dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. J. Exp. Med. 193:671–678.
3. **Chan, S. Y., C. J. Empig, F. J. Welte, R. F. Speck, A. Schmaljohn, J. F. Kreisberg, and M. A. Goldsmith.** 2001. Folate receptor-alpha is a cofactor for cellular entry by Marburg and Ebola viruses. Cell 106:117–126.
4. **Connolly, B. M., K. E. Steele, K. J. Davis, T. W. Geisbert, W. M. Kell, N. K. Jaax, and P. B. Jahrling.** 1999. Pathogenesis of experimental Ebola virus infection in guinea pigs. J. Infect. Dis. 179(Suppl. 1):S203–S217.
5. **Curtis, B. M., S. Scharnowske, and A. J. Watson.** 1992. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proc. Natl. Acad. Sci. USA 89:8356–8360.
6. **Gallaher, W. R.** 1996. Similar structural models of the transmembrane proteins of Ebola and avian sarcoma viruses. Cell 85:477–478.
7. **Geijtenbeek, T. B., D. J. Krooshoop, D. A. Bleijs, S. J. van Vliet, G. C. van Duijnhoven, V. Grabovsky, R. Alon, C. G. Figdor, and Y. van Kooyk.** 2000. DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. Nat. Immunol. 1:353–357.
8. **Geijtenbeek, T. B., D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, and Y. van Kooyk.** 2000. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell 100:587–597.
9. **Geijtenbeek, T. B., R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, G. J. Adema, Y. van Kooyk, and C. G. Figdor.** 2000. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 100:575–585.
10. **He, J., S. Choe, R. Walker, P. Di Marzio, D. O. Morgan, and N. R. Landau.** 1995. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G_2 phase of the cell cycle by inhibiting $p34^{cdc2}$ activity. J. Virol. 69:6705–6711.
11. **Lee, B., G. Leslie, E. Soilleux, U. O'Doherty, S. Baik, E. Levroney, K. Flummerfelt, W. Swiggard, N. Coleman, M. Malim, and R. W. Doms.** 2001. *cis* expression of DC-SIGN allows for more efficient entry of human and simian immunodeficiency viruses via CD4 and a coreceptor. J. Virol. 75: 12028–12038.
12. **Peters, C. J., A. Sanchez, P. E. Rollin, T. G. Ksiazek, and F. A. Murphy.** 1996. *Filoviridae*: Marburg and Ebola viruses, p. 1161–1176. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Field's virology, 3rd ed. Lippincott-Raven, Philadelphia, Pa.
13. **Pöhlmann, S., F. Baribaud, B. Lee, G. J. Leslie, M. D. Sanchez, K. Hiebenthal-Millow, J. Münch, F. Kirchhoff, and R. W. Doms.** 2001. DC-SIGN interactions with human immunodeficiency virus type 1 and 2 and simian immunodeficiency virus. J. Virol. 75:4664–4672.
14. **Relloso, M., A. Puig-Kröger, O. Muñiz, J. L. Rodríguez-Fernández, G. de la Rosa, N. Longo, J. Navarro, M. A. Muñoz-Fernández, P. Sánchez-Mateos, and A. Corbí.** 2002. DC-SIGN (CD209) expression is IL-4 dependent and is negatively regulated by IFN, TGF-β, and anti-inflammatory agents. J. Immunol. 168:2634–2643.
15. **Weitman, S. D., R. H. Lark, L. R. Coney, D. W. Fort, V. Frasca, V. R. Zurawski, Jr., and B. A. Kamen.** 1992. Distribution of the folate receptor GP38 in normal and malignant cell lines and tissues. Cancer Res. 52:3396–3401.
16. **Wool-Lewis, R. J., and P. Bates.** 1998. Characterization of Ebola virus entry by using pseudotyped viruses: identification of receptor-deficient cell lines. J. Virol. 72:3155–3160.
17. **Yang, S., R. Delgado, S. R. King, C. Woffendin, C. S. Barker, Z. Y. Yang, L. Xu, G. P. Nolan, and G. J. Nabel.** 1999. Generation of retroviral vector for clinical studies using transient transfection. Hum. Gene Ther. 10:123–132.
18. **Yang, Z., R. Delgado, L. Xu, R. F. Todd, E. G. Nabel, A. Sanchez, and G. J. Nabel.** 1998. Distinct cellular interactions of secreted and transmembrane Ebola virus glycoproteins. Science 279:1034–1037.

Cell, Vol. 100, 587–597, March 3, 2000, Copyright ©2000 by Cell Press

DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances *trans*-Infection of T Cells

Teunis B. H. Geijtenbeek,* Douglas S. Kwon,§ Ruurd Torensma,* Sandra J. van Vliet,* Gerard C. F. van Duijnhoven,* Jeena Middel,‡ Ine L. M. H. A. Cornelissen,† Hans S. L. M. Nottet,‡ Vineet N. KewalRamani,§ Dan R. Littman,§‖ Carl G. Figdor,* and Yvette van Kooyk*#
*Department of Tumor Immunology
†Department of Pathology
University Medical Center St. Radboud
Philips van Leydenlaan 25
Nijmegen 6525 EX
The Netherlands
‡Eykman-Winkler Institute of Microbiology
University Hospital Utrecht
Heidelberglaan 100
3584 CX Utrecht
The Netherlands
§Skirball Institute of BioMolecular Medicine
‖and Howard Hughes Medical Institute
New York University Medical Center
New York, New York 10016

Summary

Dendritic cells (DC) capture microorganisms that enter peripheral mucosal tissues and then migrate to secondary lymphoid organs, where they present these in antigenic form to resting T cells and thus initiate adaptive immune responses. Here, we describe the properties of a DC-specific C-type lectin, DC-SIGN, that is highly expressed on DC present in mucosal tissues and binds to the HIV-1 envelope glycoprotein gp120. DC-SIGN does not function as a receptor for viral entry into DC but instead promotes efficient infection in *trans* of cells that express CD4 and chemokine receptors. We propose that DC-SIGN efficiently captures HIV-1 in the periphery and facilitates its transport to secondary lymphoid organs rich in T cells, to enhance infection in *trans* of these target cells.

Introduction

Transmission of human immunodeficiency virus type 1 (HIV-1) infection in humans requires the dissemination of virus from sites of infection at mucosal surfaces to T cell zones in secondary lymphoid organs, where extensive viral replication occurs in $CD4^+$ T-helper cells (Fauci, 1996). These cells express both CD4 and the chemokine receptor CCR5, which together form the receptor complex required for entry by the R5 viral isolates that are prevalent early after infection (Dragic et al., 1996; Lu et al., 1997; Littman, 1998). Viruses with tropism for other chemokine receptors, particularly CXCR4, are rarely transmitted and generally appear only late in infection.

The mechanism of early viral dissemination remains vague, but based on anatomical distribution of different hematopoietic lineage cells and on in vitro infectivity studies it has been inferred that immature dendritic cells (DC) residing in the skin and at mucosal surfaces are the first cells targeted by HIV-1. DC are the most potent antigen-presenting cells in vivo (Valitutti et al., 1995; Banchereau and Steinman, 1998). Immature DC in peripheral tissues capture antigens efficiently and have the unique capacity to subsequently migrate to the T cell areas of secondary lymphoid organs. As the cells travel, they mature and alter their expression profile of cell surface molecules, including chemokine receptors, lose their ability to take up antigen, and acquire competence to attract and activate resting T cells in the lymph nodes (Adema et al., 1997; Banchereau and Steinman, 1998). HIV-1 is thought to subvert the trafficking capacity of DC to gain access to the $CD4^+$ T cell compartment in the lymphoid tissues (Grouard and Clark, 1997; Rowland-Jones, 1999; Steinman and Inaba, 1999).

Immature DC express CD4 and CCR5, albeit at levels that are considerably lower than on T cells (Granelli-Piperno et al., 1996; Rubbert et al., 1998), and they have been reported to be infectable with R5 strains of HIV-1. In contrast, immature DC do not express CXCR4 and are resistant to infection with X4 isolates of HIV-1 (Weissman et al., 1995; Blauvelt et al., 1997; Granelli-Piperno et al., 1998). Entry of HIV-1 into immature DC has also been reported to proceed through a CD4-independent mechanism (Blauvelt et al., 1997), suggesting that receptors other than CD4 could be involved. There have been conflicting reports regarding the significance of HIV-1 replication within DC (Cameron et al., 1994; Ayehunie et al., 1997; Canque et al., 1999). Although replication can be observed in some circumstances, it has also been reported that, in immature DC, replication is incomplete and that only early HIV-1 genes are transcribed.

It has been proposed that virus-infected immature DC migrate to the draining lymph nodes where they initiate both a primary antiviral immune response and a vigorous productive infection of T cells, allowing systemic distribution of HIV-1 (Cameron et al., 1992; Weissman et al., 1995; Granelli-Piperno et al., 1999). However, in a nonhuman primate model of mucosal infection with the simian immunodeficiency virus, it has been difficult to demonstrate productive infection of DC despite rapid dissemination of virus (Stahl-Hennig et al., 1999). Other efforts to model primary HIV-1 infection in vitro by exposing DC derived from skin or blood to HIV-1 have indicated that these cells are poorly infected. Nevertheless, only DC and not other leukocytes, including monocytes, macrophages, B cells, and T cells, were able to induce high levels of infection upon coculture with mitogen-activated $CD4^+$ T cells after being pulsed with HIV-1 (Cameron et al., 1992, 1992b, 1996; Weissman et al., 1995; Blauvelt et al., 1997; Granelli-Piperno et al., 1999). In an early study, Cameron et al. (1992) proposed that DC have a unique ability to "catalyze" infection of T cells with HIV but do not become infected themselves.

The mechanism by which DC capture HIV-1 and promote infection of $CD4^+$ T cells has not been elucidated,

To whom correspondence should be addressed: (e-mail: y.vankooyk@dent.kun.nl).

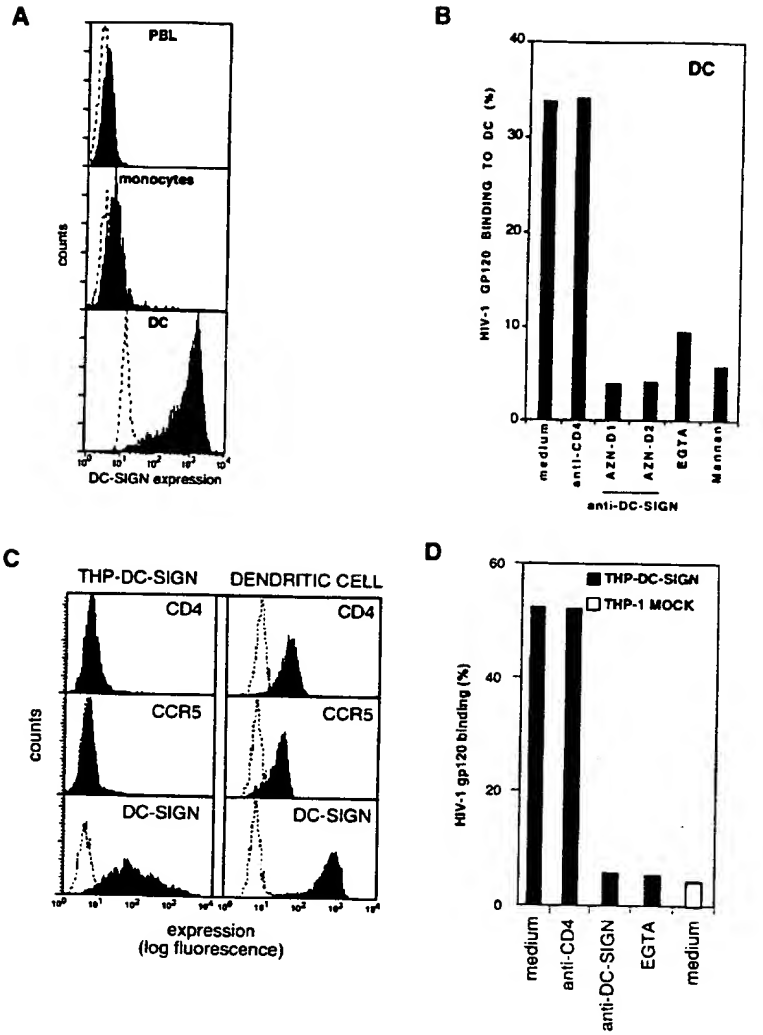


Figure 1. DC-SIGN Is a DC-Specific Receptor for HIV-1 gp120

(A) DC-SIGN is expressed specifically by DC. Immature DC, cultured from monocytes in the presence of GM-CSF and IL-4, express high levels of DC-SIGN, whereas resting peripheral blood lymphocytes and monocytes do not express DC-SIGN. Expression of DC-SIGN (AZN-D1) was determined by FACScan analysis. One representative experiment out of three is shown.

(B) DC-SIGN, but not CD4, mediates binding of HIV-1 gp120 to DC. DC were allowed to bind HIV-1 gp120-coated fluorescent beads. Adhesion was blocked by anti-DC-SIGN antibodies (20 μg/ml), mannan (20 μg/ml), and EGTA (5 mM), and not by neutralizing anti-CD4 antibodies (20 μg/ml). One representative experiment out of three is shown.

(C) Immature DC express low levels of CD4 (RPA-T4) and CCR5 (2D7/CCR5) and high levels of DC-SIGN (AZN-D1). THP-1 cells stably transfected with DC-SIGN (THP-DC-SIGN) express high levels of DC-SIGN (AZN-D1) while CD4 and CCR5 are not expressed (filled histograms). Antibodies against CD4 and DC-SIGN were isotype matched, and the appropriate isotype controls are represented by dotted lines.

(D) DC-SIGN transfectants (THP-DC-SIGN) bind HIV-1 gp120. THP-DC-SIGN and mock transfectants were allowed to bind HIV-1gp120-coated fluorescent beads. Adhesion was blocked by anti-DC-SIGN antibodies (20 μg/ml) and EGTA (5mM) and not by neutralizing anti-CD4 (RPA-T4) antibodies (20 μg/ml). One representative experiment out of three is shown.

and it has been unclear whether there is specificity in the interaction of DC with virus. In the accompanying paper, we describe the identification of a DC-specific C-type lectin, designated DC-SIGN, that binds with high affinity to ICAM-3 present on resting T cells (Geijtenbeek et al., 2000 [this issue of *Cell*]). Nucleotide sequence analysis of the cDNA indicated that this molecule is identical to a previously described HIV-1 gp120-binding C-type lectin (Curtis et al., 1992) isolated from a placental cDNA library. Here, we demonstrate that this HIV-1-binding protein, which is highly expressed on DC present at mucosal sites, specifically captures HIV-1 and promotes infection in *trans* of target cells that express CD4 and appropriate chemokine receptors. Our findings suggest that, during transmission of HIV-1, the virus initially binds to mucosal DC through DC-SIGN, allowing subsequent transport to secondary lymphoid organs and highly efficient infection of $CD4^+$ T cells by a novel *trans* infection mechanism.

Results

DC-SIGN Is a DC-Specific HIV-1-Binding Protein

DC-SIGN was recently identified as a DC-specific ICAM-3 adhesion receptor that mediates DC-T cell interactions (Geijtenbeek et al., 2000). Flow cytometric analysis of an extensive panel of hematopoietic cells with anti-DC-SIGN antibodies demonstrated that DC-SIGN is preferentially expressed on in vitro cultured DC but not on other leukocytes, such as monocytes and peripheral blood lymphocytes (PBL) (Figure 1A). Identification of DC-SIGN by peptide amino acid sequencing of the 44 kDa immunoprecipitated protein revealed it to be 100% identical in its amino acid sequence to the HIV-1 envelope glycoprotein gp120-binding C-type lectin previously isolated from a placental cDNA library (Curtis et al., 1992). To determine whether this molecule has a role in binding of HIV to DC, we used a flow cytometric adhesion assay (Geijtenbeek et al., 1999) to examine the ability of HIV-1 gp120-coated fluorescent beads to bind to immature DC (Figure 1B). The gp120-coated beads bound efficiently to the DC, and the binding was completely blocked by the anti-DC-SIGN antibodies AZN-D1 and AZN-D2. In contrast, neutralizing anti-CD4 antibodies had no effect on gp120 binding to DC. This result indicates that, although the primary HIV-1 receptor CD4 is expressed on DC (Figure 1C), HIV-1 gp120 preferentially binds to DC-SIGN. Similarly, the monocytic cell line THP-1, which lacks expression of both CD4 and CCR5, bound the gp120-coated beads after

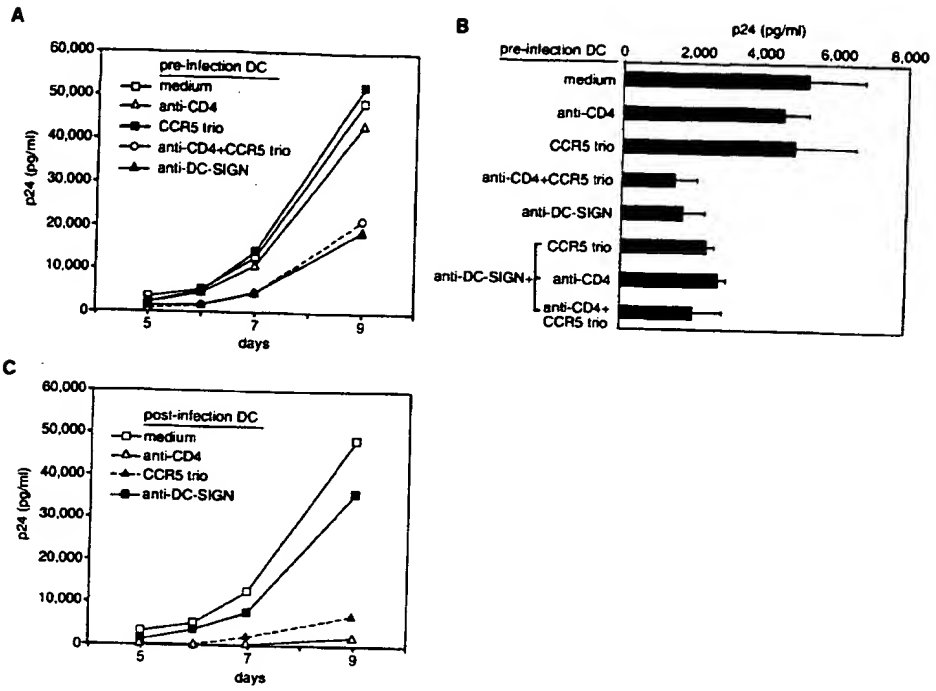


Figure 2. DC-SIGN Mediates HIV-1 Infection in a DC-T Cell Coculture

(A) Antibodies against DC-SIGN inhibit HIV-1 infection as measured in a DC-T cell coculture. DC (50×10^3) were preincubated for 20 min at room temperature with blocking mAb against CD4 (RPA-T4) or DC-SIGN (AZN-D1 and AZN-D2) (20 μg/ml) or with a combination of CCR5-specific chemokines (CCR5 trio: RANTES, MIP-1α, and MIP1β; 500 ng/ml). Preincubated immature DC were pulsed for 2 hr with HIV-1 (M-tropic HIV-1_{BaL} strain), and unbound virus particles and mAb were washed away. Subsequently, DC were cocultured with activated PBMC (50×10^3) for 9 days. Coculture supernatants were collected, and p24 antigen levels were measured by ELISA. One representative experiment out of two is shown.

(B) Inhibition of HIV-1 infection in a DC-T cell coculture by blocking DC-SIGN, CD4, and CCR5. HIV-1 replication in the DC-T cell coculture at day 5 of the experiment is described in Figure 5A. The results of day 5 are representative for days 6, 7, and 9 of DC-T cell coculture. DC were also preincubated with mAb against DC-SIGN together with anti-CD4 and CCR5-specific chemokines. p24 values represent mean ± SD of triplicate cultures. One representative experiment out of two is shown.

(C) DC-SIGN interactions with ICAM-3 are not involved in the transmission of DC-bound-HIV-1 to T cells. DC (50×10^3) were pulsed for 2 hr with HIV-1 (M-tropic HIV-1_{BaL} strain), washed, and cocultured with activated PBMC (50×10^3) for 9 days in the presence of the CCR5-specific chemokines (CCR5 trio: RANTES, MIP-1α, and MIP1β; 500 ng/ml) or mAb against CD4 (RPA-T4) and DC-SIGN (AZN-D1 and AZN-D2) (20 μg/ml). Antibodies were added post-HIV-1 infection of DC, prior to the addition of PBMC. One representative experiment out of two is shown.

it was transfected with a DC-SIGN expression vector (Figure 1C). HIV-1 gp120 binding to this cell line, THP-DC-SIGN, was also blocked by anti-DC-SIGN antibodies, but not by anti-CD4 (Figure 1D). Binding of HIV-1 gp120 to DC-SIGN expressed on DC or THP-DC-SIGN was also inhibited by the carbohydrate mannan or EGTA, consistent with previous findings (Curtis et al., 1992) and with the observation that DC-SIGN is homologous to other members of the Ca^{2+}-binding mannose-type lectins (Weis et al., 1998). Together, these results demonstrate that DC-SIGN is a specific dendritic cell surface receptor for the HIV-1 envelope glycoprotein.

DC-SIGN Is Required for Efficient HIV-1 Infection in DC-T Cell Cocultures

Because DC-SIGN is exclusively expressed on DC and has a high affinity for HIV-1 gp120, we reasoned that it might play an important role in HIV-1 infection of DC or of T cells that make contact with DC. Immature DC, which express low levels of CD4 as well as CCR5 and abundant DC-SIGN (Figure 1C), were pulsed with the R5 isolate HIV-1_{BA-L} for 2 hr, washed, and cultured in the presence of activated T cells (Figures 2A and 2B). To determine the contribution of each of these receptors in this assay system, we examined the effects of antibodies against CD4 and DC-SIGN and of a combination of three CCR5-specific chemokines (RANTES, MIP-1α, and MIP-1β). Preincubation of the immature DC with antibodies against DC-SIGN prior to infection resulted in significant inhibition of HIV-1 replication (Figure 2A). Neither anti-CD4 nor the CCR5-specific chemokines inhibited on their own, although a combination of these did block infection of DC (Figure 2A), which is probably due to efficient inhibition of the T cell infection by (un)bound anti-CD4/chemokines. Activated T cells challenged with the same viral load exhibited a weaker infection than those cultured with virus-pulsed DC (data not shown).

Since DC-SIGN binds to ICAM-3 on T cells, it is possible that antibodies against DC-SIGN could interfere with the DC-T cell interaction and thereby prevent HIV-1 transmission. To examine this possibility, antibodies against DC-SIGN were added after exposure of DC to HIV-1 but prior to the addition of activated T cells. In

this s tting, only CCR5-specific chem kines and anti-CD4 antibody strongly inhibit d HIV-1 infection of activated T cells, whil antibodies against DC-SIGN had no ffect (Figure 2C). These results thus suggest that DC-SIGN has an important function in propagation of HIV-1 in DC-T cell cocultures and that this function is related t the ability of DC-SIGN to bind to gp120 and not to its interaction with ICAM-3.

DC-SIGN Does Not Mediate HIV-1 Entry

To investigate whether DC-SIGN acts as a receptor that permits HIV-1 entry, similar to CD4 plus CCR5, we studied HIV-1 entry into 293T cells that expressed either DC-SIGN (293T-DC-SIGN) or CD4 and CCR5 (293T-CD4-CCR5). Cells were pulsed overnight with HIV_{BA-L} and washed the next day, and p24 levels were determined. There was no detectable p24 protein in the culture supernatants harvested from 293-DC-SIGN cells several days after the HIV-1 pulse, whereas the 293T-CD4-CCR5 cells were readily infected (Figure 3A).

To examine the possibility that DC-SIGN may work in conjunction with either CD4 or CCR5 to permit viral entry, we extended the studies by using HIV-1 pseudotyped with the envelope glycoprotein of the R5 isolate $HIV\text{-}1_{ADA}$. We employed a replication-defective HIV-1 genome that encoded a luciferase reporter gene, which allows a quantitative measure of the levels of single-round infection (Figure 3B) (Deng et al., 1996). Transiently transfected 293T cells expressing either CCR5 (293T-CCR5), CD4 (293T-CD4), or both (293T-CD4-CCR5), in the presence or absence of DC-SIGN, were infected with the reporter virus, and luciferase levels were determined after 2 days. As observed with replicating virus, HIV-1 entry was not detected in 293T cells that expressed only DC-SIGN (Figure 3B). No infection was observed if DC-SIGN was expressed with either CD4 or CCR5, indicating that DC-SIGN does not form a complex with these molecules to permit viral entry. In contrast, high luciferase activity was obtained after infection of 293T cells expressing both CD4 and CCR5, and expression of DC-SIGN did not contribute further to viral entry into these cells (Figure 3B). Therefore, DC-SIGN cannot substitute for CD4 or CCR5 in the process of HIV-1 entry.

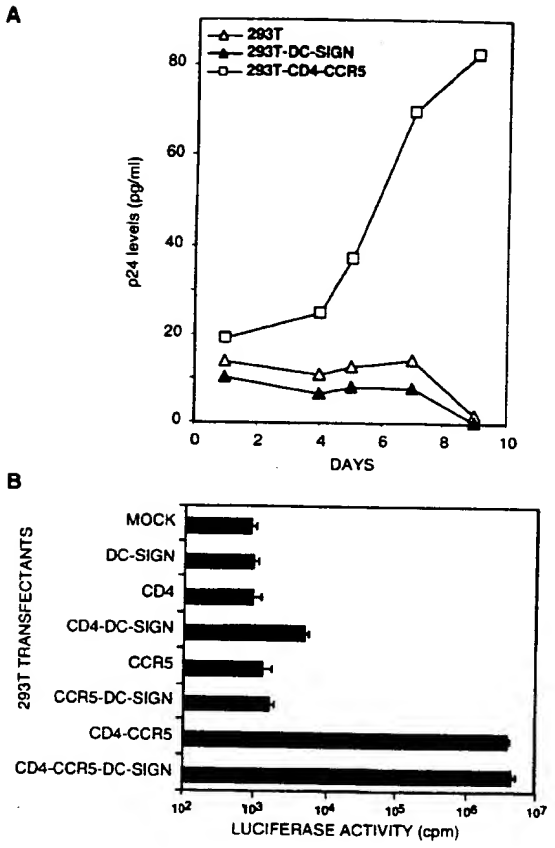


Figure 3. DC-SIGN Expressed on Target Cells Does Not Mediate HIV-1 Entry

(A) 293T cells were transfected with DC-SIGN or CD4 and CCR5 and pulsed for 2 hr with HIV-1 (CCR5-tropic $HIV\text{-}1_{BA\text{-}L}$ strain). Subsequently, cells were cultured for 9 days. Supernatants were collected, and p24 antigen levels were measured by ELISA. One representative experiment of two is shown.

(B) 293T cells and 293T cells stably expressing either CD4, CCR5, or CD4 and CCR5 were transiently transfected with DC-SIGN and subsequently infected with pseudotyped CCR5-tropic $HIV\text{-}1_{ADA}$ virus in the presence of polybrene (20 μg/ml). Luciferase activity was evaluated after 2 days. One representative experiment out of three is shown.

DC-SIGN Captures HIV-1 and Facilitates Infection of HIV-1 Permissive Cells In *trans*

Because DC-SIGN did not appear to mediate virus entry into target cells, we hypothesized that in a DC-T cell coculture (Figure 2) DC-SIGN might facilitate both capture of HIV-1 on DC, independent from CD4 and CCR5, and subsequent transmission of HIV-1 to the CD4/CCR5-positive T cells. To test this, THP-DC-SIGN transfectants, which do not express CD4 or CCR5 (Figure 1C) and which cannot be infected by HIV-1 (data not shown), were pulsed with single-round HIV-luciferase virus pseudotyped with the $HIV\text{-}1_{ADA}$ envelope glycoprotein. After washing to remove unbound virus, the cells were cocultured with CD4/CCR5-expressing 293T cells, which are permissive for HIV-1 infection, or activated T lymphocytes. THP-DC-SIGN cells were able to capture the pseudotyped virus and transmit it to the target cells that expressed the receptors required for viral entry (Figure 4A). HIV-1 capture was completely DC-SIGN dependent, as antibodies against DC-SIGN inhibited HIV-1 infection (Figure 4A), and DC-SIGN-negative parental THP-1 cells were unable to capture and transmit HIV-1 (Figures 4A and 4B). Similar to our previous findings, the DC-SIGN-mediated infection of the target cells was not due to DC-SIGN binding to ICAM-3, since 293T cells are ICAM-3 negative. These findings indicate that DC-SIGN expressed at the surface of heterologous cells can capture HIV-1 in a form that retains its capacity to subsequently infect HIV-1-permissive cells. The ability of DC-SIGN to capture and transmit HIV-1 was also observed with HIV-luciferase viruses pseudotyped with envelope glycoproteins from an additional five R5 isolates, including three primary isolates (Figure 4B), and from the X4 isolate HXB2 (data not shown).

Analysis of luciferase activity in both adherent (293T-CD4-CCR5) and nonadherent (THP-DC-SIGN) cell frac-

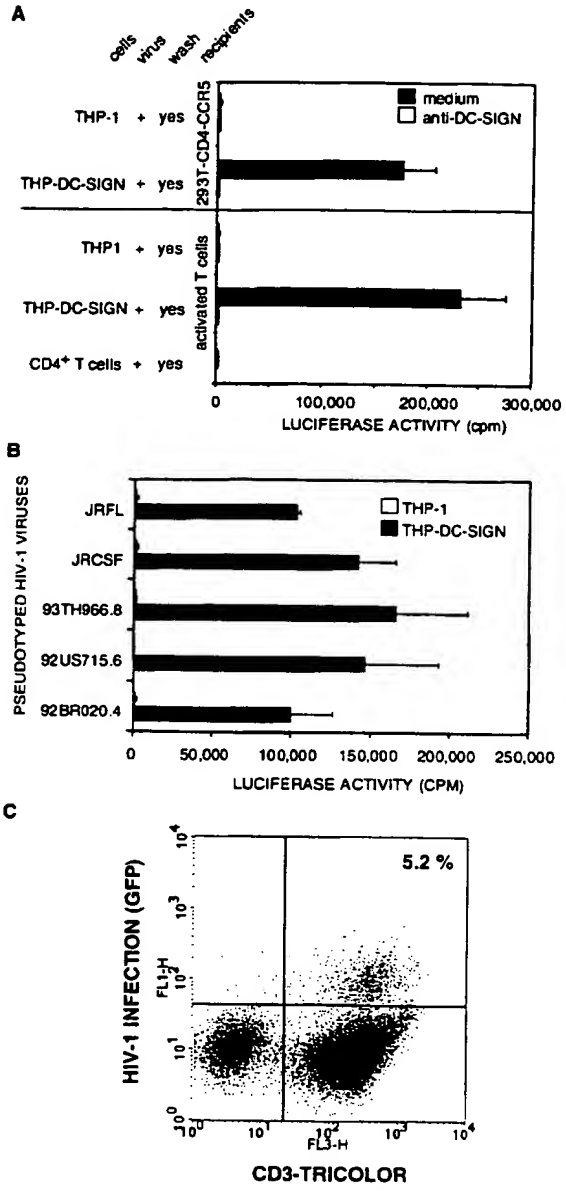


Figure 4. DC-SIGN Captures HIV-1 that Retains Infectivity for CD4+ T Cells

(A) DC-SIGN captures HIV-1 and facilitates infection of HIV-1 permissive cells in *trans*. DC-SIGN transfectants (100×10^3) were preincubated for 20 min at room temperature with blocking mAb against DC-SIGN (AZN-D1 and AZN-D2; 20 μg/ml). The THP-DC-SIGN cells were infected with HIV-luciferase virus pseudotyped with R5 strain HIV-1_{ADA} Env. Alternatively activated T cells were infected with pseudotyped HIV-1_{ADA} virus. After 2 hr at 37°C, the infected cells were extensively washed and added to either 293T-CD4-CCR5 cells or activated primary T cells (100×10^3). HIV-1 infection was determined after 2 days by measuring the luciferase activity. One representative experiment out of three is shown.

(B) DC-SIGN is able to mediate capture of HIV-1 viruses pseudotyped with M-tropic HIV-1 envelopes from different primary isolates. DC-SIGN-mediated capture was performed as described in (A) on 293T-CD4-CCR5 with HIV-luciferase viruses pseudotyped with the CCR5-specific HIV-1 envelopes from JRFL and JRCSF and from primary viruses 92US715.6, 92BR020.4, and 93TH966.8. One representative experiment out of two is shown.

(C) Activated T cells are infected by HIV-1 in the T cell/THP-DC-SIGN coculture. THP-DC-SIGN cells were incubated with HIV-eGFP viruses pseudotyed with M-tropic HIV-1_{ADA} and subsequently cocultured with activated T cells. The CD3-negative THP-DC-SIGN cells were not infected by HIV-1, whereas the CD3-positive T cells were infected. T cells, gated by staining for CD3 (tricolor), were positive for eGFP, whereas CD3-negative THP-DC-SIGN that initially captured HIV-eGFP did not express eGFP. One representative experiment out of two is shown.

tions after 2 days of coculture demonstrated that productive HIV-1 infection occured only in the HIV-1 permissive 293T-CD4-CCR5 cells (data not shown). Similarly, by using a pseudotyped HIV-1 vector with the green fluorescent protein gene in place of Nef (HIV-eGFP), we demonstrated that cells expressing CD4/CCR5 and not those expressing DC-SIGN were infected in cocultures. Thus, after coculture of virus-pulsed THP-DC-SIGN cells with T cells, only the $CD3^+$ T cells expressed virus-encoded GFP (Figure 4C).

Sexual transmission of HIV-1 is likely to require a means for small amounts of virus to gain access to cells that are permissive for viral replication. This may be achieved because of the ability of virus to interact with DC, which can capture HIV-1 and present it to the permissive cells. To mimic in vivo conditions in which HIV-1 levels are likely to be limiting, we challenged THP-1 transfectants with low titers of pseudotyped HIV-1 and subsequently cocultured these cells with HIV-1 permissive cells, without washing away unbound virus (Figure 5A). As expected, neither 293T-CD4-CCR5 cells nor activated T cells were efficiently infected with the low titers of pseudotyped HIV-1 (Figure 5A). Strikingly, when these permissive cells were challenged with an identical amount of HIV-1 in the presence of THP-DC-SIGN, but not of the parental THP-1 cells, efficient HIV-1 infection was observed in *trans* (Figure 5A). The enhancement of HIV-1 infection of primary T cells by DC-SIGN was also observed with HIV-luciferase viruses pseudotyped with five other R5 envelopes, including three from primary virus isolates (Figure 5B). These results indicate that DC-SIGN not only sequesters HIV-1 but also enhances CD4-CCR5-mediated HIV-1 entry by presentation in *trans* to the HIV-1 receptor complex. Antibodies against DC-SIGN completely inhibited infection (Figure 5A), demonstrating that the efficient enhancement of HIV-1 entry into CD4/CCR5-positive cells is DC-SIGN dependent.

DC Present in Mucosal Tissues at Sites of HIV-1 Exposure Express DC-SIGN and Are CCR5 Negative

Demonstration that cells that express DC-SIGN can capture HIV-1 and efficiently transmit the virus to other cells in *trans* suggested that DC that express this C-type lectin have a key role in viral infection in vivo. To determine whether such cells are indeed present in vivo, we performed immunohistochemical analyses of mucosal tissues that are the sites of first exposure during sexual transmission of HIV-1 (Figure 6A). DC-SIGN was expressed on DC-like cells with large and very irregular morphology that were present in the mucosal tissues, such as cervix, rectum, and uterus (Figures 6Aa, 6Ab, and 6Ac, respectively), in regions beneath the stratified

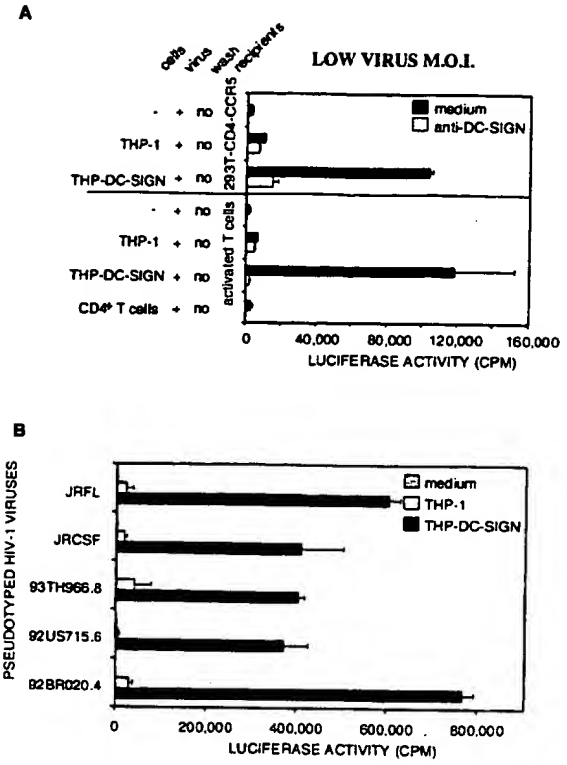


Figure 5. DC-SIGN Enhances HIV-1 Infection of T Cells by Acting In *trans*

At a low virus load, DC-SIGN in *trans* is crucial for the infection of HIV-1 permissive cells. THP-1 transfectants (100×10^3) were preincubated for 20 min at room temperature with blocking mAb against DC-SIGN (AZN-D1 and AZN-D2; 20 μg/ml). The cells were infected by low amounts of pseudotyped HIV-1_{ADA} virus (A) or other R5 isolates of HIV-1 (B), i.e., at the threshold of detection in a single round infection assay. After 1 hr at 37°C, the cell/virus suspension was directly added to either 293T-CD4-CCR5 or activated T cells (100×10^3). The infectivity was determined after 2 days by measuring the luciferase activity. One representative experiment out of two is shown.

squamous epithelium in the lamina propria. Analyses of serial sections stained for CD3, CD20, CD14, and CD68 confirmed that DC-SIGN-expressing cells were distinct from T cells, B cells, monocytes, and macrophages (data not shown). Similarly, in the accompanying paper (Geijtenbeek et al., 2000), staining of lymph nodes and skin has shown DC-restricted expression of DC-SIGN. We have also compared expression of DC-SIGN, CD4, and CCR5 on DC in the mucosa of the uterus and rectum and found in serial sections that the majority of DC-SIGN-positive DC in these tissues coexpressed CD4 but lacked CCR5 (Figure 6B). This suggests that DC present at mucosal sites, that have first contact with HIV-1 during sexual transmission, are not infected with HIV-1 through usage of CD4/CCR5. This observation is consistent with the recent demonstration that DC at sites of mucosal infection of nonhuman primates do not become infected (Stahl-Hennig et al., 1999).

DC-SIGN-Bound HIV-1 Retains Infectivity after Long-Term Culture

If HIV-1 gains access to secondary lymphoid organs by way of binding to DC, then virus would have to retain infectivity during th transport from th mucosal tissu s to the T cell z nes in draining lymph nodes. T determine if virus bound to DC-SIGN retains infectivity for a prolonged period of time, we first conducted a time-course experiment to determine the length of time that HIV-1 gp120 remains bound to DC-SIGN expressed on transfected THP-1 cells. We observed that gp120-coated beads remained bound to DC-SIGN for more than 60 hr (Figure 7A). We next investigated the length of time during which HIV-1-pulsed THP-DC-SIGN cells could retain infectious virus. The DC-SIGN-expressing transfectants were pulsed with pseudotyped HIV-1 for 4 hr and then washed extensively. The pulsed cells were subsequently placed in culture and were removed at defined intervals and cocultured with activated T cells (Figure 7B). Remarkably, after 4 days the HIV-1-pulsed cells were still able to efficiently infect target cells. In contrast, virus in the absence of DC-SIGN-positive cells lost its infectivity after 1 day. These findings support the hypothesis that limiting numbers of HIV-1 particles, captured by mucosal DC that express DC-SIGN and CD4 but not CCR5, retain infectivity during and after migration to regional lymphoid tissues. T cells, which express CD4 and CCR5, would then be productively infected due to DC-SIGN-mediated enhanced *trans* infectivity of the small numbers of HIV-1 particles (Figure 7C).

Discussion

We have identified a novel DC-specific adhesion receptor, DC-SIGN, that is identical to the high-affinity HIV-1 gp120-binding C-type lectin cloned from a human placental cDNA library (Geijtenbeek et al., 2000). We have demonstrated that DC that express both DC-SIGN and CD4 preferentially use DC-SIGN to capture HIV-1 via its high affinity for HIV-1 gp120. DC-SIGN not only efficiently recruits HIV-1 but also facilitates HIV-1 infection of $CD4^+$ T cells by a novel in *trans* mechanism. Our findings thus indicate that HIV-1 utilizes a novel receptor strategy that has not been previously described in other viral systems, and suggest that the virus exploits multiple cell surface receptor systems to ensure that it can establish a productive infection in its host organism.

DC localized in the skin and mucosal tissues such as the rectum, uterus, and cervix have been proposed to play a role in initial HIV-1 infection. DC constitute a heterogeneous population of cells that are present in minute numbers in various tissues just beneath the dermis or mucosal layer and form a first-line defense against viruses and other pathogens. DC have previously been shown to sequester HIV-1 and efficiently transmit the virus to $CD4^+$ T cells. We have demonstrated here that this property of DC can be ascribed to the ability of HIV-1 to bind specifically to these cells through the interaction of gp120 with DC-SIGN. DC thus efficiently capture HIV-1 through a specific interaction that is independent from binding of virus to CD4 and CCR5. DC-SIGN cannot mediate HIV-1 entry but rather functions as a unique HIV-1 *trans* receptor facilitating HIV-1 infection of CD4/CCR5-positive T cells (Figures 4 and 5). At low virus titer, CD4/CCR5-expressing cells are not detectably infected without the help of DC-SIGN in *trans* (Figure 5A). Conditions in which the number of HIV-1 particles is limiting are likely to resemble those

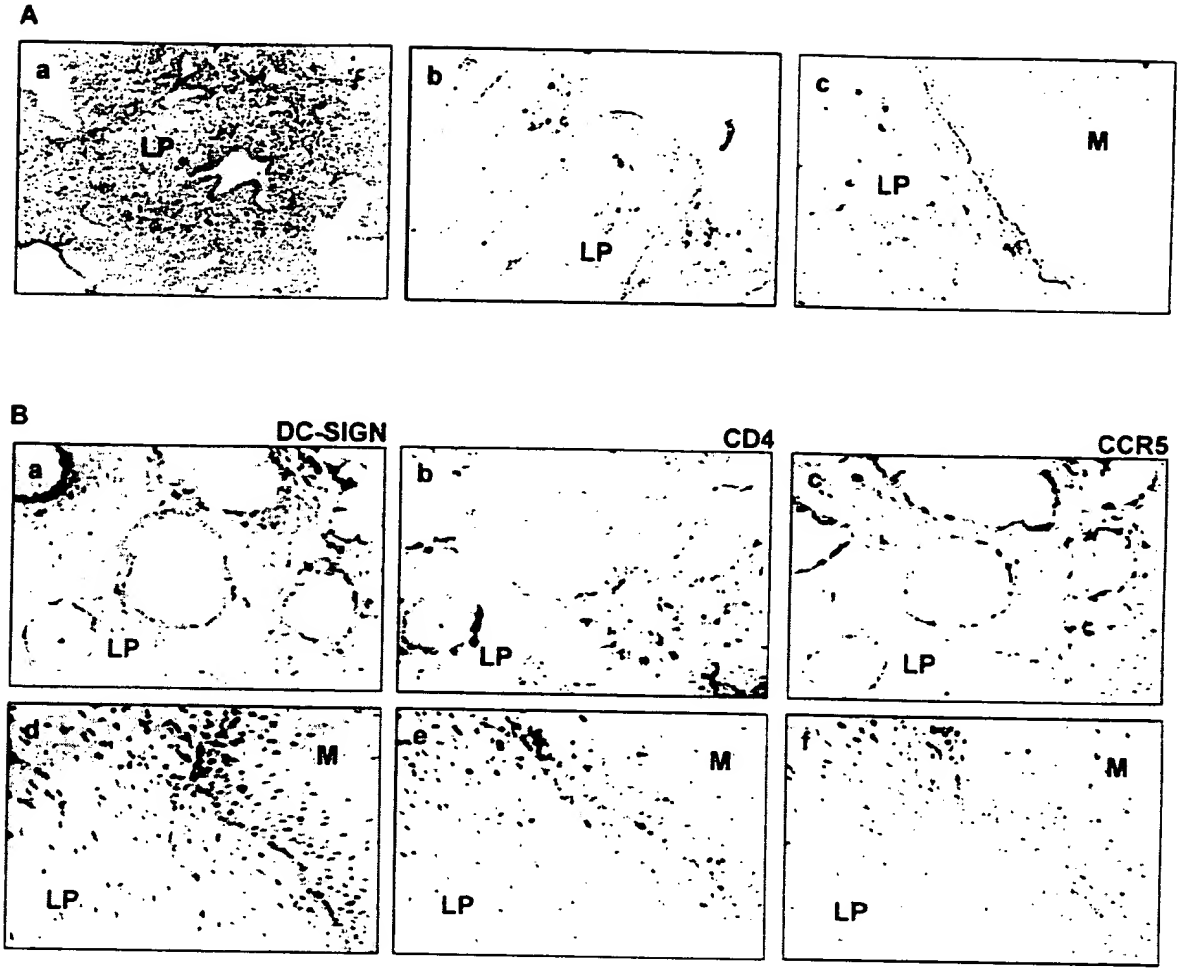


Figure 6. DC-SIGN Is Expressed on DC Present in Mucosal Tissue that Do Not Express CCR5

Immunohistochemical analysis of DC-SIGN expression on mucosal tissue sections.

(A) Different tissue sections were stained with anti-DC-SIGN mAb: cervix (a), rectum (b), and uterus (c) (original magnification, 200×) All mucosal tissues contain DC-SIGN-positive cells in the lamina propria. Staining of serial sections demonstrate that these DC-SIGN-positive cells do not express CD3, CD20, CD14, and CD68 (data not shown).

(B) Immunohistochemical staining of serial sections of rectum (a–c) and uterus (d–f) with antibodies against DC-SIGN (a and d), CD4 (b and e), or CCR5 (c and f).

found in vivo, and the results thus suggest that DC-SIGN may be required for viruses to be transmitted from mucosa to T cells that express CD4 and chemokine receptors. In addition, our studies demonstrate that virus bound to DC-SIGN is remarkably stable and can thus retain infectivity for the prolonged periods of time required for DC to traffic via lymphatics from mucosa to regional lymph nodes (Figures 7A and 7B).

Mechanism of DC-SIGN-Mediated Enhancement of HIV-1 Infectivity

The mechanisms by which HIV-1 exploits the machinery of DC and the properties of DC-SIGN to achieve efficient infection of cells that are competent for viral replication remain unclear. The process through which DC-SIGN promotes efficient infection in *trans* of cells through their CD4/chemokine receptor complex is of particular interest. Binding of the viral envelope glycoprotein to DC-SIGN may induce a conformational change that enables a more efficient interaction with CD4 and/or the chemokine receptor. As multiple conformational transitions are required before the envelope glycoprotein initiates fusion with target membranes, the binding of DC-SIGN to gp120 may facilitate or stabilize one of these transitions. Anti-gp120 antibodies that increase infectivity of viral particles have been described (Lee et al., 1997), and it is possible that DC-SIGN has a similar effect upon binding to the envelope glycoprotein. Alternatively, binding of viral particles to DC-SIGN may focus or concentrate them at the surface of the DC and may thus increase the probability that entry will occur after they bind to the receptor complex on target cells. Although the molecular mechanism has to be investigated in more detail, it is clear that DC-SIGN enhances the infection of T cells, since at low multiplicity of infection T cells are not infected in the absence of DC-SIGN.

Whether a transient quaternary complex is formed between DC-SIGN, HIV-1 Env, CD4, and CCR5 remains

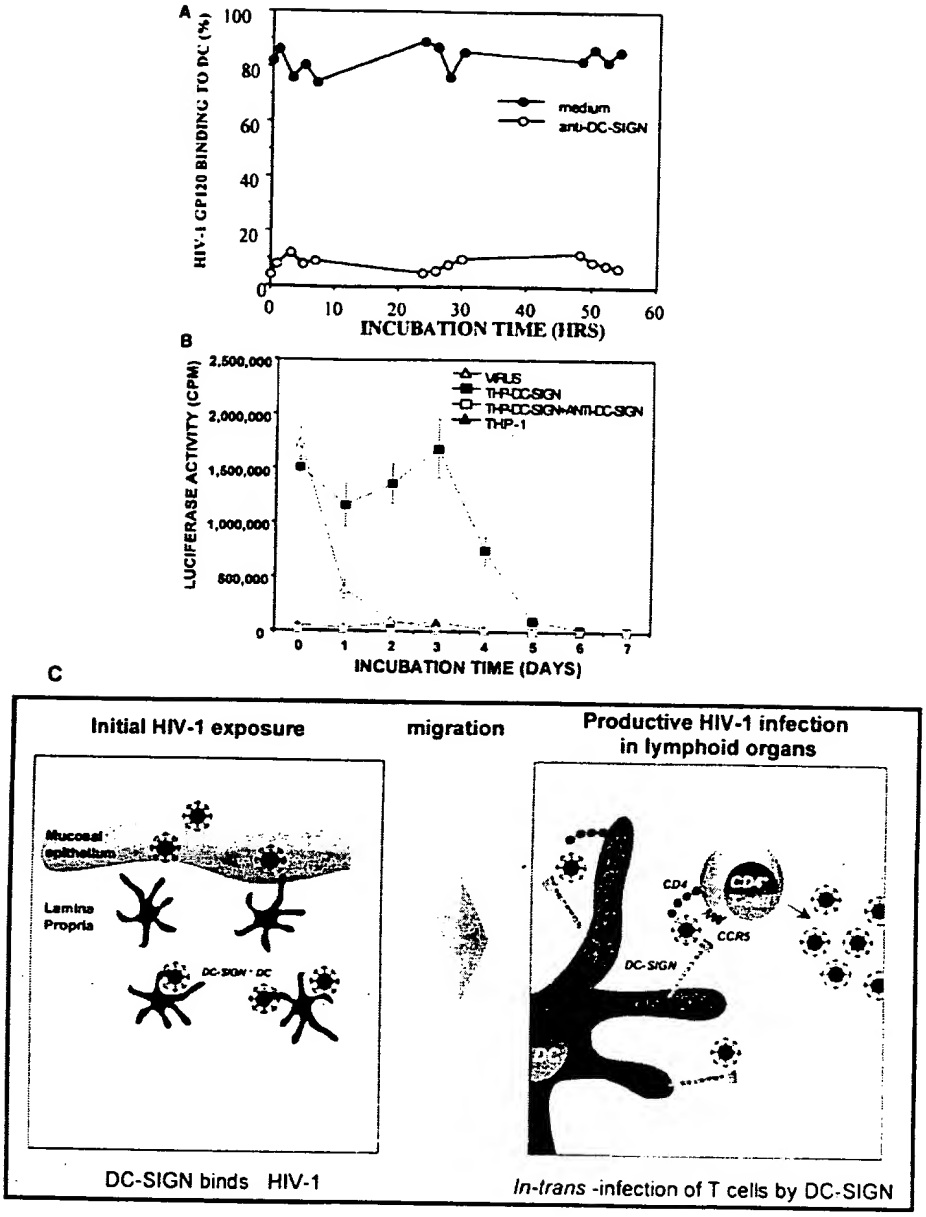


Figure 7. DC-SIGN Captures HIV-1 and Retains Long-Term Infectivity

(A) Time course of HIV-1_{MN} gp120 binding to THP-DC-SIGN. DC-SIGN-positive cells were incubated with gp120-coated beads. Beads bound for more than 60 hr as determined by FACScan analysis.

(B) DC-SIGN binds HIV-1 and retains for more than 4 days virus that infects T cells in *trans*. THP/THP-SIGN cells were pulsed for 4 hr with HIV-1 pseudotyped virus in the presence or absence of anti-DC-SIGN antibodies (AZN-D1 and AZN-D2, 20 μg/ml). After washing, the HIV-1 pulsed cells were cultured at 37°C for several days. As a control, identical amounts of virus were incubated at 37°C in medium without cells. Every day, aliquots of the HIV-1 pulsed cells were added to HIV-1 permissive 293T-CD4-CCR5 cells in order to measure infectivity. Lysates to examine luciferase activity were obtained after 2 days of coculture.

(C) Model of HIV-1 coopting DC-SIGN as a *trans* receptor after initial exposure. DC are the primary cells targeted by HIV-1 during mucosal exposure and are DC-SIGN positive. HIV-1 adheres to DC-SIGN via a high-affinity interaction, and the immature DC carrying HIV-1 migrates to the lymphoid tissues. Upon arrival, DC will cluster with T cells, and DC-SIGN enhances HIV-1 infection of T cells in *trans* leading to a productive and sustained infection.

to be determined. Elucidation of the crystal structure of a gp120-CD4 complex has revealed that most glycosylation sites within gp120 reside in a ridge that flanks the CD4-binding pocket (Kwong et al., 1998). Since mannan blocks the binding of gp120 to DC-SIGN, it is likely that this C-type lectin binds to one or more carbohydrate moieties in gp120. It remains possible, however, that the lectin domain of DC-SIGN interacts with the polypeptide

backbone f gp120. Further studies with mutant forms of gp120 and with soluble DC-SIGN may be informative in fforts t elucidat th mechanism of nhanced infectivity in *trans*.

In a separate study, we have shown that DC-SIGN binds to ICAM-3, which is expressed constitutively on the surface of T lymphocytes (Geijtenbeek et al., 2000). Enhancement of target cell infectivity by DC-SIGN-bound HIV-1 was not dependent on the presence of ICAM-3 on target cells. However, we observed that enhancement of infectivity was consistently better when target cells were T cells rather than 293-CD4-CCR5 cells. It remains possible that the efficiency of viral transmission from carrier DC to target T cells may also be enhanced by specific adhesive interactions other than DC-SIGN-ICAM-3, such as LFA-1-ICAM-1, which predominates the adhesion between DC and activated T cells (Geijtenbeek et al., 2000). Therefore, antibodies against DC-SIGN do not inhibit the DC-T cell transmission of HIV-1 postinfection (Figure 2C).

Role of DC in HIV Infection In Vivo

The only HIV-1 receptors previously known to have a role in HIV-1 entry were CD4 and a subset of the G protein–coupled chemokine receptors, including CCR5 and CXCR4. CCR5 functions as the major receptor for strains of virus previously classified as "macrophage-tropic," and only those strains that can utilize this chemokine receptor can be efficiently transmitted between individuals (Littman, 1998). Other gp120-binding receptors had been previously identified, including DC-SIGN and galactosyl ceramide (Harouse et al., 1991), but these had not been shown to be involved in viral entry. This study shows that DC-SIGN not only binds HIV-1 but can also sequester it and catalyze its entry into cells that express CD4 and chemokine receptors. Although it remains to be determined whether DC-SIGN has a significant role in HIV-1 pathogenesis in vivo, our in vitro results and the pattern of expression of the different receptors in mucosal tissues are consistent with its having a key function in the early stages of viral infection. Remarkably, our immunohistochemical analyses clearly demonstrate that CCR5 is not expressed in the lamina propria of HIV-1-related mucosal tissue (Figure 6), whereas DC-SIGN is abundantly expressed. This observation confirms and extends the findings of Hladik et al. (1999), who showed that DC present in the genital tract also lack CCR5, and strongly suggest that HIV-1 cannot infect DC present at mucosal sites.

DC-SIGN may therefore play a crucial role in initial HIV-1 exposure by mediating viral binding to DC present in mucosal tissues, rather than infection of these cells. The high level of expression of DC-SIGN on immature DC and its high affinity for gp120, which exceeds that of CD4 (Curtis et al., 1992), indicate that DC-SIGN is endowed with the ability to efficiently capture HIV-1, even when the virus is present in minute amounts. HIV-1 may subsequently exploit the migratory capacity of the DC to gain access to the T cell compartment in lymphoid tissues. DC must be activated to commence their migration, and it is hence possible that multimerization of DC-SIGN on the cell surface of DC by interaction with the multivalent virus particles may initiate this process. Interestingly, the time course experiment shows that DC-SIGN is able to captur and bind to HIV-1 for more than 4 days, after which the virus can still infect permissive c lls. This long-term preservation f HIV-1 in an infectious state would appear to allow sufficient time for it to be transported by DC trafficking from mucosal surfaces to lymphoid compartments, where virus can be transmitted (Figure 7C) (Steinman et al., 1997). Several groups have reported that DC can migrate from the periphery to draining lymph nodes within 2 days after antigen exposure or HIV-1 challenge (Barratt-Boyes et al., 1997; Stahl-Hennig et al., 1999). Viral particles have been reported within endocytic vesicles of DC. This observation suggests that DC-SIGN-bound HIV-1 may be internalized and protected during the time required for the cells to complete their journey to the regional lymph nodes. Further studies will be required to determine if viral internalization is essential for maintenance of infectivity.

Our data suggest that, after HIV-1 has been ferried by DC to the lymphoid compartment, DC-SIGN presents the bound viral particles to the CD4/CCR5 complex present on T cells and greatly enhances their entry into these cells (Figure 7C). We showed that monoclonal antibodies directed against DC-SIGN blocked productive infection occurring in the T cell cocultures with CD4/CCR5-positive monocyte-derived DC. Therefore, even in the presence of obligatory HIV-1 receptors present in *cis* on target cells, DC-SIGN functions as a *trans* receptor for HIV-1 infection of T cells and is critical in the primary cocultures. This is an important example of how a receptor can work in *trans*. Interestingly, CD4 can facilitate HIV-1 infection of CD4-negative cells that express CCR5 by a *trans* receptor mechanism, although it remains unclear whether this is an important route of infection in vivo (Speck et al., 1999). In that case, interaction of envelope glycoprotein with CD4 results in a conformational change that permits binding of the virus to CCR5 on CD4-negative cells. Together with the results presented here, these studies indicate that HIV-1 can use receptors in *trans* to facilitate infection of cells that otherwise may be difficult to infect either because of lack of proper receptors or because of their anatomical distribution relative to the sites of HIV-1 exposure.

The discovery of the role of DC-SIGN in HIV-1 infection may have significant implications for understanding the mechanism of HIV-1 transmission and for developing strategies to prevent or block viral infection. The observation that transmission of infection is confined to R5 strains of HIV-1 has remained a major enigma. In preliminary studies, we found that DC-SIGN captures and enhances infection of both X4 and R5 strains, and it is thus unlikely that preferential interaction of DC-SIGN with CCR5 would account for the restriction in tropism during transmission. Nevertheless, the demonstration that uninfected DC contribute to the process of viral entry raises the possibility that the requirement for CCR5 utilization may reflect a requirement for interaction of multiple cell types. The inhibition of HIV-1 infection observed in the presence of anti-DC-SIGN antibodies suggests that interfering with the gp120-DC-SIGN interaction either during the capture phase of DC in the mucosa or during DC/T cell interactions in lymphoid organs could inhibit dissemination of the virus. Small molecule inhibitors, potentially carbohydrate-based, that block

th ability of gp120 to bind to DC-SIGN may b effective in prophylaxis or therapeutic interv ntion. Vaccine strategi s aimed at liciting mucosal antibodies that inhibit gp120 binding to DC-SIGN may also be efficacious in preventing early establishment of infection. The efficacy of gp120 vaccines has been measured as a function of th levels of neutralizing antibodies that inhibit HIV entry through CD4 and CCR5. Our results thus suggest that levels of antibodies that block virus binding to DC-SIGN or the DC-SIGN-mediated enhancement of infection may also be predictive of protection.

Experimental Procedures

Antibodies

The following mAb were used: 2D7 (anti-CCR5; Becton Dickinson and Co., Oxnard, CA) and CD4 (RPA-T4; PharMingen, San Diego, CA). Anti-DC-SIGN mAb AZN-D1 and AZN-D2 were obtained by screening hybridoma supernatants of human DC-immunized BALB/c mice for the ability to block adhesion of DC to ICAM-3, as measured by the fluorescent bead adhesion assay.

Cells

Immature DC were cultured as previously described (Geijtenbeek et al., 2000). Stable THP-1 transfectants expressing DC-SIGN were generated by transfection of THP-1 cells with pRc/CMV-DC-SIGN by electroporation similarly as described (Lub et al., 1997).

Fluorescent Bead Adhesion Assay

Carboxylate-modified TransFluorSpheres (488/645 nm, 1.0 μm; Molecular Probes, Eugene, OR) were coated with M-tropic HIV-1_{MN} envelope glycoprotein gp120 similarly as was described for ICAM-1 beads (Geijtenbeek et al., 1999). Streptavidin-coated beads were incubated with biotinylated F(ab')2 fragment rabbit anti-sheep IgG (6 μg/ml; Jackson Immunoresearch) followed by an overnight incubation with sheep-anti-gp120 antibody D7324 (Aalto Bio Reagents Ltd., Dublin, Ireland) at 4°C. The beads were washed and incubated with 250 ng/ml purified HIV-1 gp120 (provided by Immunodiagnostics, Inc., through the NIH AIDS Research and Reference Reagent Program) overnight at 4°C. The fluorescent beads adhesion assay was performed as described by Geijtenbeek et al. (1999).

HIV-1 Infection of Both DC and DC-SIGN Transfectants

The M-tropic strain HIV-1_{Ba-L} was grown to high titer in monocyte-derived macrophages (MDM). Seven days after titration of the virus stock on MDM, $TCID_{50}$ was determined with a p24 antigen ELISA (Diagnostics Pasteur, Marnes la Coquette, France) and estimated as 10^4/ml. DC (50×10^3) preincubated with mAb against DC-SIGN (AZN-D1 and AZN-D2) or CD4 (RPA-T4) (20 μg/ml) or a combination of CCR5-specific chemokines (RANTES, MIP-1α, MIP-1β; each 500 ng/ml) for 20 min at room temperature were pulsed for 2 hr with HIV-1_{Ba-L} (at a multiplicity of infection of 10^3 infectious units per 10^5 cells), washed, and cocultured with activated PBMC (50×10^3). No DC-T cell syncytium formation was observed. The postinfection experiment was performed similarly except that the mAb or chemokines were added after the washing step of the HIV-1 pulse, together with the activated PBMC. Culture supernatants were collected at day 5, 6, 7, and 9 after DC-T cell coculture and p24 antigen levels, as a measure of HIV-1 production were determined by a p24 antigen ELISA. PBMC were activated by culturing them in the presence of IL-2 (10 U/ml) and PHA (10 μg/ml) for 2 days.

Pseudotyped viral stocks were generated by calcium-phosphate transfections of 293T cells with the proviral plasmid pNL-Luc-E^-R^- (containing a luciferase reporter gene) or the proviral pHIV-eGFP (containing a GFP reporter gene) and expression plasmids for ADA, JRFL, and JRCSF gp160 envelopes. The isolation, identification, and construction of the plasmids encoding the primary virus envelopes from 92US715.6, 92BR020.4, and 93TH966.8 has been previously described (Bjorndal et al., 1997). Viral stocks were evaluated by limiting dilution on 293T-CD4-CCR5 cells. HIV-1 pseudotyped with murine leukemia virus (MLV), amphotropic Env, and vesicular stomatitis virus glycoprotein (VSV-G) were used to ensure target cell viability.

Immunohistochemical analyses were performed as described previously (Geijtenbeek et al., 2000).

Acknowledgments

This work was supported by the Dutch Cancer Society (the NKB) (grant number 96-1358), the Netherlands Organization for Scientific Research (the NWO) (grant number 901-09-244), and the Technology Foundation (the STW) (grant number NGN.4187). V. N. K. is a fellow of the Damon Runyon-Walter Winchell Foundation, and D. R. L. is an investigator of the Howard Hughes Medical Institute and was supported in this project by the National Institutes of Health grant AI33856. We also thank H. Jacobs for the generation of THP-DC-SIGN cells. We are grateful to Dr. D. J. Ruiter and Dr. G. N. Muijen for their advice on the immunohistochemical analyses and Dr. H. Schuitemaker and Dr. N. Kootstra for their helpful discussions.

Received November 30, 1999; revised January 19, 2000.

References

Adema, G.J., Hartgers, F., Verstraten, R., de Vries, E., Marland, G., Menon, S., Foster, J., Xu, Y., Nooyen, P., McClanahan, T., et al. (1997). A dendritic-cell-derived C-C chemokine that preferentially attracts naive T cells. Nature *387*, 713–717.

Ayehunie, S., Garcia-Zepeda, E.A., Hoxie, J.A., Horuk, R., Kupper, T.S., Luster, A.D., and Ruprecht, R.M. (1997). Human immunodeficiency virus-1 entry into purified blood dendritic cells through CC and CXC chemokine coreceptors. Blood *90*, 1379–1386.

Banchereau, J., and Steinman, R.M. (1998). Dendritic cells and the control of immunity. Nature *392*, 245–252.

Barratt-Boyes, S.M., Watkins, S.C., and Finn, O.J. (1997). In vivo migration of dendritic cells differentiated in vitro: a chimpanzee model. J. Immunol. *158*, 4543–4547.

Bjorndal, A., Deng, H., Jansson, M., Fiore, J.R., Colognesi, C., Karlsson, A., Albert, J., Scarlatti, G., Littman, D.R., and Fenyo, E.M. (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. J. Virol. *71*, 7478–7487.

Blauvelt, A., Asada, H., Saville, M.W., Klaus-Kovtun, V., Altman, D.J., Yarchoan, R., and Katz, S.I. (1997). Productive infection of dendritic cells by HIV-1 and their ability to capture virus are mediated through separate pathways. J. Clin. Invest. *100*, 2043–2053.

Cameron, P.U., Freudenthal, P.S., Barker, J.M., Gezelter, S., Inaba, K., and Steinman, R.M. (1992). Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to $CD4^+$ T cells. Science *257*, 383–387.

Cameron, P.U., Lowe, M.G., Crowe, S.M., O'Doherty, U., Pope, M., Gezelter, S., and Steinman, R.M. (1994). Susceptibility of dendritic cells to HIV-1 infection in vitro. J. Leukoc. Biol. *56*, 257–265.

Cameron, P., Pope, M., Granelli-Piperno, A., and Steinman, R.M. (1996). Dendritic cells and the replication of HIV-1. J. Leukoc. Biol. *59*, 158–171.

Canque, B., Bakri, Y., Camus, S., Yagello, M., Benjouad, A., and Gluckman, J.C. (1999). The susceptibility to X4 and R5 human immunodeficiency virus-1 strains of dendritic cells derived in vitro from CD34(+) hematopoietic progenitor cells is primarily determined by their maturation stage. Blood *93*, 3866–3875.

Chan, D.C., and Kim, P.S. (1998). HIV entry and its inhibition. Cell *93*, 681–684.

Curtis, B.M., Scharnowske, S., and Watson, A.J. (1992). Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proc. Natl. Acad. Sci. USA *89*, 8356–8360.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., et al. (1996). Identification of a major co-receptor for primary isolates of HIV-1. Nature *381*, 661–666.

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., et al. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature *381*, 667–673.

Fauci, A.S. (1996). Host factors and the pathogenesis of HIV-induced disease. Nature *384*, 529–534.

Geijtenbeek, T.B., van Kooyk, Y., van Vliet, S.J., Renes, M.H., Raymakers, R.A., and Figdor, C.G. (1999). High frequency of adhesion defects in B-lineage acute lymphoblastic leukemia. Blood *94*, 754–764.

Geijtenbeek, T.B., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Adema, G.J., van Kooyk, Y., and Figdor, C.G. (2000). Identification of DC-SIGN, a novel dendritic cell–specific ICAM-3 receptor that supports primary immune responses. Cell *100*, 575–585.

Granelli-Piperno, A., Moser, B., Pope, M., Chen, D., Wei, Y., Isdell, F., O'Doherty, U., Paxton, W., Koup, R., Mojsov, S., et al. (1996). Efficient interaction of HIV-1 with purified dendritic cells via multiple chemokine coreceptors. J. Exp. Med. *184*, 2433–2438.

Granelli-Piperno, A., Delgado, E., Finkel, V., Paxton, W., and Steinman, R.M. (1998). Immature dendritic cells selectively replicate macrophagetropic (M-tropic) human immunodeficiency virus type 1, while mature cells efficiently transmit both M- and T-tropic virus to T cells. J. Virol. *72*, 2733–2737.

Granelli-Piperno, A., Finkel, V., Delgado, E., and Steinman, R.M. (1999). Virus replication begins in dendritic cells during the transmission of HIV-1 from mature dendritic cells to T cells. Curr. Biol. *9*, 21–29.

Grouard, G., and Clark, E.A. (1997). Role of dendritic and follicular dendritic cells in HIV infection and pathogenesis. Curr. Opin. Immunol. *9*, 563–567.

Harouse, J.M., Bhat, S., Spitalnik, S.L., Laughlin, M., Stefano, K., Silberberg, D.H., and Gonzalez-Scarano, F. (1991). Inhibition of entry of HIV-1 in neural cell lines by antibodies against galactosyl ceramide. Science *253*, 320–323.

Hladik, F., Lentz, G., Akridge, R.E., Peterson, G., Kelley, H., McElroy, A., and McElrath, M.J. (1999). Dendritic cell-T-cell interactions support coreceptor-independent human immunodeficiency virus type 1 transmission in the human genital tract. J. Virol. *73*, 5833–5842.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature *393*, 648–659.

Lee, S., Peden, K., Dimitrov, D.S., Broder, C.C., Manischewitz, J., Denisova, G., Gershoni, J.M., and Golding, H (1997). Enhancement of human immunodeficiency virus type 1 envelope-mediated fusion by a CD4-gp120 complex-specific monoclonal antibody. J. Virol. *71*, 6037–6043.

Littman, D.R. (1998). Chemokine receptors: keys to AIDS pathogenesis? Cell *93*, 677–680.

Lu, Z., Berson, J.F., Chen, Y., Turner, J.D., Zhang, T., Sharron, M., Jenks, M.H., Wang, Z., Kim, J., Rucker, J., et al. (1997). Evolution of HIV-1 coreceptor usage through interactions with distinct CCR5 and CXCR4 domains. Proc. Natl. Acad. Sci. USA *94*, 6426–6431.

Lub, M., van Vliet, S.J., Oomen, S.P., Pieters, R.A., Robinson, M., Figdor, C.G., and van Kooyk, Y. (1997). Cytoplasmic tails of beta 1, beta 2, and beta 7 integrins differentially regulate LFA-1 function in K562 cells. Mol. Biol. Cell *8*, 719–728.

Rowland-Jones, S.L. (1999). HIV: the deadly passenger in dendritic cells. Curr. Biol. *9*, R248–R250.

Rubbert, A., Combadiere, C., Ostrowski, M., Arthos, J., Dybul, M., Machado, E., Cohn, M.A., Hoxie, J.A., Murphy, P.M., Fauci, A.S., et al. (1998). Dendritic cells express multiple chemokine receptors used as coreceptors for HIV entry. J. Immunol. *160*, 3933–3941.

Speck, R.F., Esser, U., Penn, M.L., Eckstein, D.A., Pulliam, L., Chan, S.Y., and Goldsmith, M.A. (1999). A trans-receptor mechanism for infection of CD4-negative cells by human immunodeficiency virus type 1. Curr. Biol. *9*, 547–550.

Stahl-Hennig, C., Steinman, R.M., Tenner-Racz, K., Pope, M., Stolte, N., Matz-Rensing, K., Grobschupff, G., Raschdorff, B., Hunsmann, G., and Racz, P. (1999). Rapid infection of oral mucosal-associated lymphoid tissue with simian immunodeficiency virus. Science *285*, 1261–1265.

Steinman, R.M., and Inaba, K. (1999). Myeloid dendritic cells. J. Leukoc. Biol. *66*, 205–208.

Steinman, R.M., Pack, M., and Inaba, K. (1997). Dendritic cells in the T-cell areas of lymphoid organs. Immunol. Rev. *156*, 25–37.

Valitutti, S., Muller, S., Cella, M., Padovan, E., and Lanzavecchia, A. (1995). Serial triggering of many T-cell receptors by a few peptide-MHC complexes. Nature *375*, 148–151.

Weis, W.I., Taylor, M.E., and Drickamer, K. (1998). The C-type lectin superfamily in the immune system. Immunol. Rev. *163*, 19–34.

Weissman, D., Li, Y., Ananworanich, J., Zhou, L.J., Adelsberger, J., Tedder, T.F., Baseler, M., and Fauci, A.S. (1995). Three populations of cells with dendritic morphology exist in peripheral blood, only one of which is infectable with human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA *92*, 826–830.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A61K 31/70, 31/715

A2

(11) International Publication Number: WO 93/01820

(43) International Publication Date: 4 February 1993 (04.02.93)

(21) International Application Number: PCT/US92/05985

(22) International Filing Date: 16 July 1992 (16.07.92)

(30) Priority data:
731,214 16 July 1991 (16.07.91) US

(71) Applicant: BRISTOL-MYERS SQUIBB COMPANY [US/US]; 345 Park Avenue, New York, NY 10154 (US).

(72) Inventor: CURTIS, Benson, M. ; 1520 N.W. Woodbine Way, Seattle, WA 98177 (US).

(74) Agents: POOR, Brian, W.; Bristol-Myers Squibb Company, 3005 First Avenue, Seattle, WA 98121 (US) et al.

(81) Designated States: AU, CA, FI, JP, KR, NO, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

(57) Abstract

A specific nonCD4-gp120 receptor has been isolated which has specific binding affinity for the gp120 surface protein f human immunodeficiency virus (HIV). Methods of treating HIV infection of CD4 negative cells, such as colon and brain, are disclosed together with methods of detecting HIV and diagnostic kits.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

TECHNICAL FIELD OF THE INVENTION

The present invention is directed to a non-CD4 cell surface receptor for gp120. This gp120 receptor (gp120r) has been isolated and cloned and is utilized in the present invention in methods and kits for the inhibition and detection of HIV infection.

BACKGROUND OF THE INVENTION

Two types of human retroviruses have been identified, leukemia viruses and AIDS-related viruses. The primary targets of the human retroviruses are T lymphocytes and cells of the central nervous system. All human retroviruses are transmitted by intimate contact, blood contamination, and infection in utero or after birth by milk. It is likely that all human retroviruses originated in Africa and that they encountered the human species via interspecies infection, possibly from African green monkeys or a related species. The human retroviruses first discovered, Human T Lymphotropic Virus Type 1 (HTLV-1) and Human T Lymphotropic Virus Type II (HTLV-II), have a preferential tropism for T4 cells and some T8 cells, share significant sequence homology, and are mainly associated with T cell leukemias and lymphomas. The other group of human retroviruses, generally called Human Immunodeficiency Viruses (HIV), is discussed in greater detail below. There are two major differences between the two types of human retroviruses: (1) there is substantial genomic variability among various HIV isolates, whereas the genomes of HTLV-I and HTLV-II are stable; and (2) HIV entered human populations much more recently than HTLV-I or HTLV-II.

The human immunodeficiency virus (HIV) is a cytopathic retrovirus and the causative agent of the acquired immunodeficiency syndrome (AIDS). Two forms of HIV have now been identified. The prototype virus, HIV-1, previously termed lymphadenopathy-associated virus (LAV) and Human T Lymphotropic Virus Type III (HTLV-III), is responsible for the vast majority of reported AIDS cases worldwide. Another retrovirus, HIV-2, has been isolated primarily from West African patients with AIDS and is pathogenically related to HIV-1. On the genetic level, HIV-2 is actually more closely related to the simian immunodeficiency virus (SIV), a retrovirus infecting monkeys.

Over half of the people that have contracted AIDS in the United States have already died. As many as three million persons in this country may be asymptomatic carriers of

HIV and are capable of transmitting the virus. It had been estimated in 1986 that 270,000 cases of AIDS will have occurred in the United States by 1991 (U.S. Public Health Service, (1986), Public Health Rep. 101:341). The mortality rate from AIDS is disturbingly high, exceeding 80% within three years of diagnosis and possibly reaching 100% over a longer period.

Worldwide, the AIDS epidemic may involve some five to ten million presently infected persons. Particularly troublesome are statistics from the African continent where millions of individuals are believed infected with HIV, deaths range in the hundreds of thousands, and heterosexual transmission predominates. To date, there is neither a known cure for AIDS nor an effective vaccine against HIV infection.

HIV is a member of the nontransforming, cytopathic lentivirus family of retroviruses. HIV causes a typically fatal disease characterized by severe immunodeficiency or neurodegenerative disease, or both. The primary basis for HIV induced immunosuppression is the depletion of the helper/inducer subset of T lymphocytes expressing the CD4 molecule (T4 or $CD4^+$ cells), which serves as a high affinity cell surface receptor for the virus. T4 lymphocytes are involved directly or indirectly in the induction of nearly every immunologic function in the body, and their depletion results in susceptibility to a wide range of opportunistic infections and neoplasms.

In addition to the T4 lymphocyte, other cells expressing the CD4 molecule are targets of HIV infection, especially monocyte-macrophages. HIV infection also results in serious B cell abnormalities including polyclonal activation, hypergammaglobulinemia, elevated levels of circulating immune complexes, and autoantibodies. A decreased number of functional natural killer (NK) cells have also been observed in AIDS patients.

Infection of $CD4^+$ cells is initiated by the interaction of the CD4 molecule with the major HIV envelope glycoprotein gp120, an event which is followed by internalization and uncoating of the virion, transcription of genomic RNA to DNA by virus-encoded reverse transcriptase, and integration of the resulting proviral DNA into host cell chromosomal DNA. Also, unintegrated proviral DNA accumulates in large amounts within infected cells and is probably a significant factor in HIV cytopathology (Shaw et al., (1984) Science 226:1165).

The depletion of $CD4^+$ T cells appears to contribute significantly to the immunosuppression associated with AIDS. A primary cytopathic effect of the virus in vitro is HIV-induced syncytium formation. CD4, through its interaction with gp120 plays an important role in syncytium formation. However, it has been observed that molecules on the cell surface of uninfected cells other than CD4 are also involved in HIV-induced cell fusion (Hildreth et al. (1989) Science 244:1075-1078).

Infection by HIV produces, in addition to AIDS, a set of neuropsychiatric disorders which are called the AIDS dementia complex (ADC) (Price et al., (1988) 239:586-592).

The symptoms of ADC include cognitive impairment, apathy and motor dysfunctions, and may affect as many as 90% of AIDS victims. The underlying cause of ADC appears to be the death of brain cells and HIV-1 can be isolated from the brains of infected individuals (Ho et al, (1987) N. Eng. J. Med. 317:278-286).

An early study suggested that the cellular attachment site for HIV in brain might be CD4 (Pert et al., (1986) Proc. Natl. Acad. Sci. USA 83:9254-9258) but attempts to replicate these findings were not successful (Kozlowski et al., (1989) Neurosci. Abstr. 15:671). It now appears unlikely that the CD4 antigen is involved in the infection of brain-derived cells by HIV. Susceptibility of brain cells to infection with HIV-1 does not correlate with the level of expression of CD4 (Chang-Mayer et al., (1987) Proc. Natl. Acad. Sci. USA 84:3526-3530; Srinivasan et al., (1988) Arch. Virol. 98:135-141), and infection of brain-derived cells by HIV-1 is not blocked by anti-CD4 antibodies (Clapham et al., (1989) Nature 337:368-370; Li et al., (1990) J. Virol. 64:1383-1387).

The present invention demonstrates the presence of a non-CD4 receptor for gp120 and a method for the inhibition of HIV infection of cells such as brain and muscle which do not express high levels of CD4.

SUMMARY OF THE INVENTION

Many cells that are susceptible to HIV infection appear to bind gp120 through a non-CD4 surface protein. The present invention has identified this non-CD4 gp120 receptor (gp120r) and has recombinantly expressed and characterized gp120r.

In this invention a specific non-CD4 gp120r has been isolated which has specific binding activity for gp120 present on Human Immunodeficiency Virus-1 (HIV). This gp120r has a molecular weight of about 45, 000 daltons, contains about 400 amino acid residues and is characterized by a Kd for gp120 of about 1.3 nM to about 2.0 nM. The binding of gp120 to gp120r is inhibited by specific carbohydrates, such as mannose and fucose, plant lectins such as concanavalin A and specific antibiotics, such as pradimicin A.

In one embodiment of the present invention, a cDNA molecule that transcribes an mRNA encoding for gp120r is cloned and expressed to produce gp120r. The DNA is selected from a gene library obtained from tissue such as placenta, brain, muscle and colon.

A method of inhibiting HIV infection of mammalian cells, such as brain, muscle and neural cells, is contemplated by the present invention. In this method, cells are contacted with an effective amount of an appropriate inhibitor of gp120r binding for a time

period sufficient to significantly inhibit the binding of HIV to the non-CD4 protein, gp120r. Specific inhibitors of gp120r binding include mannose carbohydrates, fucose carbohydrates, plant lectins, and antibiotics such as pradimicin A.

The gp120r of the present invention can also be utilized in a method and a kit for the detection of the presence of HIV in a fluid sample. In this method, the binding of HIV to gp120r is detected by an indicating means such as a labelled antibody capable of binding to the HIV-gp120r reaction product. It is also contemplated that the gp120r can be affixed to a solid matrix to form a solid support that is useful in this method and/or kit.

DESCRIPTION OF THE FIGURES

In the drawings:

FIGURE 1 illustrates expression cloning of the gp120r cDNA and comparison to CD4.

A: Autoradiography of gp120 binding to gp120r and CD4 expressed in COS cells. A-F [^{125}I]vgp120; A, gp120r; B, gp120r with G17-2; C, gp120r with 200 nM unlabelled bgp120; D, CD4; E, CD4 with G17-2; F, CD4 with bgp120. G-L [^{125}I]ngp120; G, gp102r; H, gp120r with 110.1; I, gp120r with bgp120; J, CD4: k, CD4 with 110.1; L, CD4 with bgp120.

B: Inhibition of [^{125}I]vgp120 binding to gp120r and CD4. A-F gp120r and G-L CD4. A+G, HIV antisera (1:20; Trimar); B+H, D-galactose (100 mM); C+I, D-mannose (100 mM); D+J, L-fucose (100 mM); E+K, Concanavalin A (1 mg/ml); F+L, pradimicin A (100 μg/ml).

C: gp120r binding of HIV. A, HIV; B, HIV with 200 nM bgp120.

FIGURE 2 illustrates the characterization of the gp120r.

A: Scatchard analysis of [^{125}I]gp120 binding. ▲ - ▲, vgp120 binding to placenta, Kd 1.3 nM, B_{max} 19 fmol/mg protein; ■ with μg/ml G17-2; ● - ●, vgp120 binding to gp120r COS cells, Kd 1.7 nM, B_{max} 150,000 receptors/cell (R/C); ○, ngp120, Kd 1.8 nM, 149,000 R/C.

B: Inhibition of [^{125}I]gp120 binding to gp120r COS cells. Open symbols ngp120, filled symbols vgp120. The relative values were the same with both forms of gp120. Mannan expressed as mg/ml. ❑, mannan (IC_{50} 6 μg/ml); , L-fucose (K_i 6 mM); Δ, α-methyl D-mannoside (K_i 15 mM), ○, D-mannose (K_i 23 mM); ◇, N-acetylglucosamine (K_i 70 mM), ■, EGTA (K_i 0.3 mM).

C: Internalization of gp120 by gp120r COS cells. Points represent the mean of two experiments with vgp120 and ngp120. - , suface; O - O internal.

D: Placenta control sera; 2, placenta HIV sera; 3, gp120r COS control sera; 4, gp120r COS HIV sera.

E: Northern blot of gp120r expression. Polyadenylated (A^+); 2, placenta; 3, thymus; 4+12, forebrain; 5, skeletal muscle; 6, heart; 7, liver; 8, kidney; 9, colon; 10 medulla; 11, cerebellum; 13, T cell (CEM; 16 μg A^+) 14, B cell (TS-1; 16 μg A^+); 15, macrophage (U937; 8 μg A^+); 16, cervical carcinoma (HeLa; 16 μg A^+). The different apparent size of the ~5 kb band is an artifact of displacement by 28S rRNA.

FIGURE 3 illustrates the sequence analysis of the gp120r.

A: Nucleotide and deduced protein sequence of gp120r cDNA.

B: Hydropathicity plot of the gp120r. The predicted transmembrane segment and the start of the eight amphipathic repeats are indicated by arrows.

C: Aminoacid alignment of the gp120r C-type lectin domain.

DESCRIPTION OF PREFERRED EMBODIMENTS

HIV infection of brain and muscle cell lines is not blocked by soluble CD4 or anti-CD4 antibodies (Clapham, P.R. et al., (1989) Nature 337:368-370; Harouse, J.M. et al., (1989) J. Virol. 63:2527-2533; Weber, J. et al., (1989) J. Gen. Virol. 70:2653-2660). This is consistent with the existence of a second gp120 receptor. Binding studies indicated that human placenta was another source for a non-CD4 gp120 receptor, and a cDNA for a second gp120 receptor (gp120r) was isolated by the present invention from a placental library. The gp120r has a higher binding affinity for gp120 than CD4. Sequence analysis revealed homology to membrane associated C-type lectins, and inhibition studies have shown that the receptor binds gp120 through a mannose or fucose containing carbohydrate. The gp120r rapidly internalizes gp120, and is expressed in placenta, thymus, muscle, and colon. These results, when considered with previous studies on the role of gp120 carbohydrate in HIV infection (Lifson, J. et al., (1986) J. Exp. Med. 164:2101-2106; Ezekowitz, R.A.B. et al., (1989) J. Exp. Med. 169:185-196; Larkin M. et al., (1989) AIDS 3: 793-798; Tanabe-Tochikura A. et al., (1990) Virology 176:473-476), suggest a potential role for the gp120r in HIV infection or pathology.

The present invention demonstrates that the gp120r participates in cellular binding of HIV by a non-CD4 pathway in muscle and brain, as well as, facilitating virus attachment in CD4 positive cell types. It is likely that the gp120r plays a significant role in

transplacental transport of HIV (Zacher, V. et al., (1991) J. Virol. 65:2102-2107) and colon infection (Barnett, S.W. et al. (1991) Virol. 182:802-809). Gp120 produces an increase in intracellular calcium in rat retinal ganglion cells (Dreyer, E.B. et al., (1990) Science 248:364-367) suggesting that the gp120r or a homologous protein may have signaling functions in the nervous system disrupted by gp120 leading to HIV neurotoxicity.

In the present invention, a new non-CD4 binding protein, or receptor, for gp120 was isolated. The HIV surface protein gp120 was found to bind to a receptor on human placental membranes that was not blocked by antibodies directed against CD4, such as G17-2 and OKT4a, and which interfere with gp120 binding to CD4. A cDNA encoding this receptor was isolated from a placental cDNA library in a mammalian expression vector (pCDM8). The gene products were expressed in COS cells and were screened by ^{125}I-labelled gp120 binding. From a pool of 90,000 cDNA molecules, a single clone was isolated that encoded a protein which bound gp120, even in the presence of concentrations of anti-CD4 antibody (G17-2) which completely blocked gp120 binding to CD4.

Sequence studies were carried out and indicated that the 1.5 kilobase cDNA clone encoded a previously unknown member of a family of Type II membrane proteins with an extracellular C type lectin domain.

The cloned gp120r of the present invention binds gp120 with an affinity (Kd) of about 1 to 2 nM, which is considerably greater than the affinity of CD4 for gp120 (about Kd = 4 nM).

The binding of gp120 to gp120r is not blocked by polyclonal HIV antisera, but is inhibited by mannose carbohydrates, fucose carbohydrates, plant lectins such as concanavalin A and pradimicin A antibiotics. Other sugars such as N-acetyl-d-glucosamine and galactose are less potent inhibitors.

The gp120r is expressed on many mammalian cells which do not exhibit high levels of CD4, such as placenta, skeletal muscle, brain, and mucosal cells. Other tissue and cells displaying gp120r include colon, thymus, heart, T cells, B cells and macrophages. The distribution of tissue having gp120r parallels that for binding of gp120 which is not blocked by CD4 antibodies, and for HIV infection which is not neutralized by soluble CD4. This observation suggests a role for gp120r in viral infection.

In gp120r expressing transfected COS cells, gp120 is rapidly internalized following binding to gp120r. This binding and internalization of gp120 is inhibited by compounds such as mannan, concanavalin A and pradimicin A.

In the present invention a cDNA which encodes gp120 was isolated and cloned. A DNA molecule of the present invention corresponds to a complementary DNA molecule which transcribes a messenger RNA (mRNA) molecule which, when translated, encodes gp120r. The cDNA molecules were obtained by reverse-transcribing mRNA molecules

isolated from mammalian tissue such as placenta, colon, brain or thymus. The transcription and cloning of cDNA molecules and isolation of gene products are techniques well known in the art and, for example, are described in Sambrook et al., "Molecular Cloning: A Laboratory Manual", 2d edition, Cold Spring Harbor Lab., Cold Spring Harbor, NY (1989), which is incorporated herein by reference.

As used herein, the phrases "physiologically tolerable" and "pharmaceutically acceptable" refer to molecular entities and compositions that do not produce an allergic or similar untoward reaction, such as gastric upset, dizziness and the like, when administered to a mammal. The physiologically tolerable carrier may take a wide variety of forms depending upon the preparation desired for administration and the intended route of administration.

A carrier is a material useful for administering the active compound and must be "acceptable" in the sense of being compatible with the other ingredients of the composition and not deleterious to the recipient thereof.

The pharmaceutical compositions are prepared by any of the methods well known in the art of pharmacy all of which involve bringing into association the active compound and the carrier therefor.

For therapeutic use, the agent utilized in the present invention can be administered in the form of conventional pharmaceutical compositions. Such compositions can be formulated so as to be suitable for oral or parenteral administration, or as suppositories. In these compositions, the agent is typically dissolved or dispersed in a physiologically tolerable carrier.

As an example, the compounds of the present invention can be utilized in liquid compositions such as sterile suspensions or solutions, or as isotonic preparations containing suitable preservatives. Particularly well suited for the present purposes are injectable media constituted by aqueous injectable isotonic and sterile saline or glucose solutions. Additional liquid forms in which the present compounds may be incorporated for administration include flavored emulsions with edible oils such as cottonseed oil, sesame oil, coconut oil, peanut oil, and the like, as well as elixirs and similar pharmaceutical vehicles.

The present agents can also be administered in the form of liposomes. As is known in the art, liposomes are generally derived from phospholipids or other lipid substances. Liposomes are formed by mono- or multi-lamellar hydrated liquid crystals that are dispersed in an aqueous medium. Any non-toxic, physiologically acceptable and metabolizable lipid capable of forming liposomes can be used. The present compositions in liposome form can contain, in addition to the agent of the present invention, stabilizers,

preservatives, expedients, and the like. The preferred lipids are the phospholipids and the phosphatidyl cholines (lecithins), both natural and synthetic.

Methods to form liposomes are known in the art. See, for example, Prescott, Ed., "Methods in Cell Biology", Volume XIV, Academic Press, New York, N.Y. (1976) p 33 et seq.

The present compounds can also be used in compositions such as tablets or pills, preferably containing a unit dose of the compound. To this end, the agent (active ingredient) is mixed with conventional tabletting ingredients such as corn starch, lactose, sucrose, sorbitol, talc, stearic acid, magnesium stearate, dicalcium phosphate, gums or similar materials as non-toxic, physiologically tolerable carriers. The tablets or pills of the present compositions can be laminated or otherwise compounded to provide unit dosage forms affording prolonged or delayed action.

It should be understood that in addition to the aforementioned carrier ingredients the pharmaceutical formulation described herein can include, as appropriate, one or more additional carrier ingredients such as diluents, buffers, flavoring agents, binders, surface active agents, thickeners, lubricants, preservatives (including antioxidants) and the like, and substances included for the purpose of rendering the formulation isotonic with the blood of the intended recipient.

The tablets or pills can also be provided with an enteric layer in the form of an envelope that serves to resist disintegration in the stomach and permits the active ingredient to pass intact into the duodenum or to be delayed in release. A variety of materials can be used for such enteric layers or coatings, including polymeric acids or mixtures of such acids with such materials as shellac, shellac and cetyl alcohol, cellulose acetate, and the like. A particularly suitable enteric coating comprises a styrene-maleic acid copolymer together with known materials that contribute to the enteric properties of the coating.

A method of inhibiting HIV infection of mammalian cells is disclosed in the present invention. A pharmaceutical composition containing a compound which effectively inhibits the binding of gp120r to HIV, is contacted with cells either in vitro or in vivo for a time period sufficient to significantly inhibit the binding of HIV to the cell surface.

Compounds effective in this method include mannose carbohydrates, fucose carbohydrates, plant lectins and pradimicin A antibiotics. Specifically preferred compounds are mannose, fucose, mannan, concanavalin A and pradimicin A. The pharmaceutical composition of the present invention includes a compound which effectively inhibits gp120r binding to HIV and may also include a physiologically tolerable carrier.

The method of the present invention is preferably utilized to inhibit HIV infection of placental, brain, muscle, neural and colon cells.

A diagnostic method is also described in the present invention for detecting the presence, and preferably the amount, of HIV present in a fluid sample by producing a reaction product containing HIV bound to gp120r. Those skilled in the art will recognize that there are well known clinical diagnostic procedures that can be utilized for the formulation and detection of such reaction products. Thus, while exemplary assay methods are described herein, the invention is not intended to be so limited.

Various heterogeneous and homogeneous assay protocols can be employed for detecting the presence, and preferably the amount, of HIV in a fluid sample. For example, the present invention contemplates a method for assaying a sample, such as a body fluid, for the presence of HIV comprising the steps of:

(a) admixing a fluid sample with gp120r, either in solution or affixed to a solid matrix;

(b) maintaining the admixture for a predetermined time period such as about 10 minutes to about 16 - 20 hours and under biological assay conditions at a temperature of about 4°C to about 45°C that is sufficient for any HIV present in the sample to react with (bind) the gp120r to form a reaction product; and

(c) determining the presence of any reaction product that is formed, and thereby the presence of any HIV in the admixture.

Preferably, the fluid sample is a body fluid sample, such as blood, plasma, serum, urine, saliva, semen or cerebrospinal fluid (CSF).

The determination of the presence of a reaction product, either directly or indirectly, can be accomplished by assay techniques well known in the art such as by the use of an indicating or labelling means, as discussed hereinbelow. In a preferred embodiment, a labelled indicating means, such as a fluorescein-labelled antibody, is capable of binding to the gp120r present in the reaction product to form a labelled complex. Determining the presence of the labelled complex provides an assay for the presence of HIV in the sample. In particularly preferred embodiments, the amount of labelled indicating means bound as part of the complex is determined, and thereby the amount of HIV present in the sample is determined. When that amount is zero, no HIV is present in the sample, within the limits of detection. Methods for assaying the presence and amount of a labelled indicating means depend on the label used, such labels and assay methods being well known in the art.

In a preferred embodiment, the gp120r is affixed on a solid matrix to form a solid phase support. In that embodiment, the assay is heterogeneous, solid/liquid phase assay and, as such, has its own preferred manipulations. For example, following admixing of a liquid sample with a solid support containing gp120r affixed thereto, the admixture is

maintained under biological assay conditions for a time period sufficient for any HIV present in the sample to bind to gp120r and form a solid phase bound reaction product. The solid and liquid phases are then separated to remove any material in the sample that did not react with the solid support, such as by rinsing. This removes any material present in the sample that could interfere with the detection of the reaction product.

A labelled indicating means is then admixed with the separated solid phase in an aqueous medium to form a solid/liquid phase labelling-reaction admixture which is maintained for a time period sufficient for the indicating means to bind to the solid bound reaction product forming a labelled complex. The solid phase is then separated from the liquid phase, rinsed and the presence, and preferably amount, of the indicating means present is determined.

As used herein, the term "biological assay conditions" refers to parameters that maintain the biological activity of the molecules and organisms in the present invention, and include a temperature range of about 4°C to about 45°C, a pH value range of about 5 to about 9, and an ionic strength varying from that of distilled water to that of about one molar sodium chloride. Methods for optimizing such conditions are well known in the art.

As used herein, the term "about" refers to a range of values both greater than and/or less than the listed value by 10% or less. For example, a temperature of about 20° C will include temperature values of from 18° C to 22° C.

As used herein, the term "corresponds", and its various grammatical modifications, means "is similar or in agreement with".

A diagnostic system in kit form for assaying a fluid sample for the presence of HIV is also contemplated by the present invention. Such a kit includes, in an amount sufficient for at least one assay, gp120r as a packaged reagent, together with instructions for use. An indicating means capable of detecting or signalling the presence of a reaction product formed between gp120r and HIV may also be present in the kit as a separately packaged reagent.

As used herein, the term "instructions for use" typically includes a tangible expression describing the reagent concentration or at least one assay method parameter such as the relative amounts of reagent and sample to be admixed, maintenance time periods for admixtures, temperature, buffer conditions and the like.

The packaging materials discussed herein in relation to diagnostic systems are those customarily utilized. Such materials include glass and plastic (e.g. polyethylene, polypropylene and polycarbonate) bottles, vials, plastic and plastic-foil laminated envelopes and the like.

As used herein, the term "package" refers to a solid material such as glass, plastic, paper, foil and the like capable of holding within fixed limits the gp120r, and preferably

also a detection means. In one embodiment, the package can contain a microtiter plate well to which microgram quantities of gp120r have been operatively affixed, ie., linked so as to be capable of reacting with and bind HIV and/or gp120.

As used herein, the terms "label" "indicating means" and "labelled indicating means", in their various grammatical forms refer to single atoms and molecules that are either directly or indirectly involved in the production of a detectable signal to indicate or detect the presence of a reaction product. Such labels are themselves well known in clinical diagnostic chemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel methods and/or systems.

The indicating means can be a fluorescent labelling agent that chemically binds to antibodies or protein antigens without denaturing them to form a fluorochrome (dye) that is a useful immunofluorescent tracer. Suitable fluorescent labelling agents are fluorochrome, such as fluorescein isocyanate (FIC), fluorescein isothiocyanate (FITC), 5-dimethylamine-1-naphthalene sulfonyl chloride (DANSC), tetramethylrhodamine isocyanate (TRITC), lissamine and the like. Immunofluorescence analysis techniques are well known in the art, and for example, is described in DeLuca, "Immunofluorescence Analysis" in Immunofluorescence Analysis, Marchalonis et al., (1982) eds., John Wiley & Sons, Ltd., pp. 189-231, which is incorporated herein by reference.

Other preferred indicating means are colorimetric agents and enzymes, such as horseradish peroxidase, glucose oxidase or the like, linked as described above, as well as radioactive elements, preferably an element that produces gamma ray emissions. Elements which emit gamma rays, such as ^{124}I, ^{125}I, ^{128}I, ^{132}I, and ^{51}Cr represent one class of radioactive indicating groups. Another group of useful labelling means are those elements such as ^{11}C, ^{18}F, ^{15}O and ^{13}N which emit positrons. The positrons so emitted produce gamma rays upon interaction with electrons present.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

EXAMPLE 1

Cloning and Isolation of Non-CD4 Gp140 Receptor Protein

Human placental membranes were found to be able to bind vaccinia derived recombinant gp120 (vgp120) with a Kd of 1.3 nM. At nM (concentrations) of gp120 none of this binding was inhibited by an antibody (G17-2) which has been reported to efficiently block gp120 binding to CD4 (Linsley et al. (1988) J. Virol. 62:3695-3702), as shown in

FIGURE 2A. Approximately 50 - 90% of the total placental gp120 binding was not due to CD4.

A placental cDNA library was obtained in the mammalian expression vector pCDM8 and was screened. A cDNA was isolated which expressed protein that exhibited high affinity binding for vgp120 in the presence of G17-2.

This protein, designated as gp120 receptor (gp120r), also bound native gp120 (ngp120), and the binding component was precipitated in the presence of an antibody directed against gp120.

EXAMPLE 2

Characterization

The binding of radiolabelled gp120 to gp120r expressed in COS-7 cells was studied. Pools of 90,000 cDNA molecules, obtained from a placental pCDM8 library, were transfected by electroporation into COS-7 cells. Cells which expressed gp120r on the surface was identified by screening with either 1 nM of ^{125}I-labelled vgp120 (^{125}I-vgp120) or ^{125}I-ngp120 by the method described in Kozlowski et al., (1990) Antivir. Chem. Chemother. 1:175-182, incorporated herein by reference. The results of binding studies utilizing the transfected COS-7 cells are shown in FIGURE 1.

Binding of labelled gp120 (1 nM) to the cells was carried out following a 1 hour preincubation of the cells or GP120 at 22°C with one or more of the following: anti-CD4 antibody G17-2 (5 ug/ml), baculovirus-derived gp120 (bgp120, American Biotechnologies, 200 nM), anti-gp120 monoclonal antibody 110.1 (25 μg/ml), D-mannose (100 mM), D-galactose (100 mM), L-fucose (100 mM), concanavalin A (1 mg/ml) or pradimicin A (100 ug/ml). The cells were monitored after autoradiography (3 days). The results seen in FIGURES 1 (A and B) illustrate that gp120 binding to the gp120r expressed on the cells was blocked by excess bgp120, mannose, fucose, pradimicin A, Concanavalin A, and preincubation with antibody 110.1 but not by CD4, antibody G17-2, galactose, or HIV antisera. Studies were also carried out on gp120 binding to CD4 expressing COS cells, transfected with π H3MCD4 by the method of Peterson et al. (1988) Cell 54:65-72.

Control studies of the binding of ^{125}I-labelled psoralen-UV inactivated HIV-BRU to the gp120r expressing COS-7 cells demonstrated binding of HIV to gp120r and blockage by excess bgp120 (FIGURE 1C). A tabular compilation quantitating the amount of bound material to the cells in FIGURE 1 is shown in Table. 1

TABLE 1

FIGURE 1	LABELLED MATERIAL	COMPETITION WITH	CPM BOUND X 10^{-3} GP120R	CD4
A	vgp120+	—	60	20
		G17-2	60	5
		bgp120	6	4
B	vgp120+	HIV antisera	63	3
		D-galactose	50	20
		D-mannose	6	20
		L-fucose	6	20
		concanavalin A	8	6
		pradimicin A	8	6
		OKT4A	60	5
		N-acetylgalactosamine	60	20
		N-acetylglucosamine	30	20
		mannan	6	20
		mannose-6-phosphate	60	20
		sialic acid	60	20
		human IgE	60	20
C	HIV-BRU+	—	8	4
		bgp120	2	2

Scatchard plots of gp120 binding to placental membranes and to COS cells expressing the gp120r were carried out in the presence and absence of a 200 fold excess of bgp120 or ngp120. The results, shown in FIGURE 2A, demonstrate a specific binding of vgp120 to gp120r with a Kd of 1.7 nM $\pm$ 0.4 (n=4) and of ngp120 to gp120r with Kd of 1.8 nM $\pm$ 0.2 (n=4), with 150,000 and 149,000 receptors per cell, respectively. Concurrent analysis of gp120 binding to CD4 expressed on COS cells gave a Kd of 4-5 nM in agreement with previous reports (Linsley, P.S. et al. (1988) J. Virol 62:3695-3702; Schnittman, et al. (1988) J. Immunol. 141:4181-4186). Calculations from the association and dissociation rate constants gave a similar comparative result. The expressed gp120r has a relative molecular mass (Mr) of ~48,500 and a protein of similar size was also partially purified from placental membranes (FIGURE 2D).

The placental membranes and COS cells were surface iodinated, and treated with 1 nM unlabelled vgp120, then washed with Blotto RPMI, 5% BSA, 1% Non-fat dry milk, 0.2% sodium azide solubilized in Triton X-100 (1% in PBS with a protein inhibitor cocktail, PMSF, Pepstatin A, orthophenathroline and leupeptin) and immunoprecipitated

with HIV or control human sera, according to the method described in Curtis et al. (1990) J. Immunol. 144:1295-1303.

Northern analysis of the expression of the gp120r RNA indicated a major species of ~5 kb and a minor species of ~1.7 kb which may represent an alternatively processed transcript and is more consistent with the size of the gp120r cDNA. RNA was denatured, separated in an agarose gel, transferred to nitrocellulose, hybridized to gp120r cDNA and autoradiographed for 3 days.

Expression of gp120r RNA was highest in colon followed by thymus, placenta, heart, skeletal muscle, and was not detected in liver or kidney. Low levels of expression in brain, T cell, B cell, and macrophage (FIGURE 2E) require verification by polymerase chain reaction (PCR). Full length CD4 RNA was highest in thymus, T cell, and macrophage followed by placenta and colon (not shown).

The gp120r cDNA encodes a protein of 404 amino acids with a calculated Mr of 45,775 (FIGURE 3A).

Sequencing of both strands of gp120r cDNA was carried out by the dideoxy chain termination method. The nucleotide sequence preceeding the first ATG agrees with the Kozak consensus. The predicted cytoplasmic domain has a similar length and shows some sequence homology to other type II membrane protein C-type lectins (Spiess, M. (1990) Biochemistry 29:10009-10018). The membrane spanning sequence is underlined and was predicted in part by homology to related sequences in FIGURE 3C. The potential N-linked glycosylation site is marked by an asterisk. The start of the seven complete and eighth partial tandem repeats are indicated (R1-R8). The consensus repeat sequence is IYQELT(R/Q) LKAAVGELPEKSKLQE. The beginning of the lectin domains is also indicated (L). No signal sequence was apparent but instead demonstrated homology to a family of Type II membrane proteins which utilize a ~20 residue hydrophobic stop-transfer sequence for membrane translocation. The "positive inside rule" (von Heijne, G. et al. (1988) Eur. J. Biochem. 174:671-678) for the sequence within fifteen residues of the transmembrane region predicts a cytoplasmic amino terminus in agreement with the homology to membrane associated C-type lectins with similar membrane orientation (FIGURE 3C) (Spiess, M. (1990) Biochemistry 29:10009-10018). This region, Met 1 to Ala 76, represents the first domain of the gp120r sequence.

The second domain (Ile 77 to Val 249) consists of tandem repeats of nearly identical sequence (FIGURE 3A). This region was predicted to consist of a series of amphipathic α-helices interrupted by ß-turns. Circular Dichroism spectra in 40% trifluoroethanol of a consensus repeat peptide beginning with the ß-turn, PEKSKLQEIYQELTQLKAAVGEL (single-letter amino-acid code), demonstrated an all α-helical structure (not shown). Homology to other repeat domains suggested three possible

tertiary structures, (1) antiparallel helix bundles, (2) a multimeric parallel helix bundle, and (3) a membrane pore with a hydrophobic exterior and a negatively charged interior. The first two models would function as spacers to separate the lectin domain from the membrane, while the third could generate a transmembrane signal after ligand binding.

The third domain (Cys 253 to Ala 404) is homologous to the other known C-type lectins which are type II membrane proteins (FIGURE 3C). With the exception of the IgEr, these lectins bind terminal D-galactose and D-N-acetylgalactosamine of glycoproteins (Spiess, M. (1990) Biochemistry 29: 10009-10018).

The most closely related sequences were the group of Type II membrane protein C-type lectins: Chick hepatic lectin (CHL) (Drickamer, K.J. (1981) Biol. Chem. 256:5827-5839), low affinity IgE receptor (IgEr) (Kikutani, H. et al. (1986) Cell 47: 657-665), the asialoglycoptorein receptors (human H1 and H2 (Spiess, M. et al. (1985) Proc. Natl. Acad. Sci. USA 82:6465-6569) are shown), and the rat Kupffer cell receptor (Hoyle, G.W. et al. (1988) J. Biol. Chem. 263:7487-7492). The most similar mannose binding lectin was one of the eight carbohydrate recognition domains of the human macrophage mannose receptor (Mannr) (Taylor, M.E. et al. (1990) J. Biol. Chem. 265:12156-12162; Ezekowitz, R.A.B. et al. (1990) J. Exp. Med. 172:1785-1794). Residues identical to the gp120r are boxed. ALIGN scores indicate significant sequence similarity if greater than 3.0. The complete gp120r sequence was most homologous to the Kupffer cell receptor which has a similar tandem repeat (Hoyle, G.W. et al. (1988) J. Biol. Chem. 263:7487-7492).

The inability to crosslink gp120 to the non-CD4 sites on placenta and brain cell lines (not shown) was consistent with an interaction of the gp120r with carbohydrate, and polyclonal HIV antisera added to gp120 blocked binding to CD4 but not to the gp120r (FIGURE 1B). Galactose and N-acetylgalactosamine did not block gp120 binding, but mannose and fucose completely blocked binding to the gp120r without an effect on CD4 (FIGURE 1B). Inhibition by a series of sugars is shown in FIGURE 2B. Human IgE (10 μg/ml), sialic acid (100 mM), and mannose-6-phosphate (100 mM) had no effect on binding to the gp120r. The three forms of gp120 used have different oligosaccaride structures. Bgp120 contains only high mannose structures (Hsieh, P. et al. (1984) J. Biol. Chem. 259:2375-2382). Vgp120 has equal proportions of high mannose and complex (Mizuochi, T. et al. (1988) Biochem. J. 254:599-603) similar to ngp120 which has a greater structural diversity in the complex chains (Geyer, H. et al. (1988) J. Biol. Chem. 263:11760-11767; Mizuochi, T. et al. (1990) J. Biol. Chem. 265:8519-8524). The affinity of the gp120r for all three forms was similar (FIGURE 2A) suggesting that the terminal mannose of high mannose chains are the primary determinants of binding. As expected for a C-type lectin the gp120r required calcium and binding was blocked by

EGTA (FIGURE 2B). The gp120r carbohydrate specificity is more closely related to serum mannose-binding proteins and to the Mr 175,000 mannose-specific endocytosis receptor found in macrophages and placenta (Taylor, M.E. et al. (1990) J. Biol. Chem. 265:12156-12162; Ezekowitz, R.A.B. et al. (1990) J. Exp. Med. 172:1785-1794) (FIGURE 3C). Low (1 nM) concentrations of gp120 did not purify a Mr 175,000 band from placental membranes (FIGURE 2D) consistent with a reported concentration of 150-300 nM for gp120 saturation of the macrophage receptor (Larkin, M. et al. (1989) AIDS 3, 793-798).

The importance of gp120 carbohydrate in HIV infection has been suggested by the ability of plant lectins (Lifson, J. et al. (1986) E. J. Exp. Med. 164:2101-2106) and serum mannose-binding protein (Ezekowitz, R.A.B. et al. (1989) J. Exp. Med. 169:185-196) to block infection, and a proposed role for the macrophage endocytosis receptor in viral attachment (Larking M. et al. (1989) AIDS 3, 793-798). Concanavalin A treatment of gp120 blocked binding to the gp120r and CD4 (FIGURE 1B), consistent with a steric hindrance of receptor interaction. The antibiotic pradimicin A blocks HIV infection of CD4 positive T cells and this inhibitory effect is prevented by mannan and EGTA (Tanabe-Tochikura. A. et al. (1990) Virology 176:476-473). Pradamicin blocked gp120 binding to the gp120r and CD4, while mannan and EGTA only inhibited binding to the gp120r (FIGURE2B). Mannan inhibited ~10% of high affinity (nM) gp120 binding to T cells and macrophages, consistent with gp120r expression (FIGURE 2E), suggesting that in addition to CD4 the gp120r may be important for HIV binding and infection. The observation the the gp120r rapidly internalized its bound ligand gp120 (FIGURE 2C), and also binds radiolabelled HIV in a gp120 dependent fashion (FIGURE 1C) also support this conclusion.

The foregoing description and Examples are intended as illustrative of the present invention, but not as limiting. Numerous variations and modifications may be effected without departing from the true spirit and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Curtis, Benson

(ii) TITLE OF INVENTION: INHIBITION OF NON-CD4 MEDIATED HIV INFECTION

(iii) NUMBER OF SEQUENCES: 9

(iv) CORRESPONDENCE ADDRESS:

(A)	ADDRESSEE:	Bristol-Myers Squibb Company
(B)	STREET:	3005 First Avenue
(C)	CITY:	Seattle
(D)	STATE:	Washington
(E)	COUNTRY:	USA
(F)	ZIP:	98121

(v) COMPUTER READABLE FORM:

(A)	MEDIUM TYPE:	Floppy disk
(B)	COMPUTER:	IBM PC compatible
(C)	OPERATING SYSTEM:	PC-DOS/MS-DOS
(D)	SOFTWARE:	PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A)	APPLICATION NUMBER:	US UNKNOWN
(B)	FILING DATE:	11-JUL-1991
(C)	CLASSIFICATION:	

(viii) ATTORNEY/AGENT INFORMATION:

(A)	NAME:	Sorrentino, Joseph M.
(B)	REGISTRATION NUMBER:	32,598
(C)	REFERENCE/DOCKET NUMBER:	ON0086-

(ix) TELECOMMUNICATION INFORMATION:

(A)	TELEPHONE:	(206) 728-4800
(B)	TELEFAX:	(206) 448-4775

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	1312 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	double
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Human immunodeficiency virus type 1

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 42..1253

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CTAAAGCAGG AGTTCTGGAC ACTGGGGGAG AGTGGGGTGA C ATG AGT GAC TCC          53
                                                 Met Ser Asp Ser
                                                   1

AAG GAA CCA AGA CTG CAG CAG CTG GGC CTC CTG GAG GAG GAA CAG CTG        101
Lys Glu Pro Arg Leu Gln Gln Leu Gly Leu Leu Glu Glu Glu Gln Leu
  5                  10                  15                  20

AGA GGC CTT GGA TTC CGA CAG ACT CGA GGA TAC AAG AGC TTA GCA GGG        149
Arg Gly Leu Gly Phe Arg Gln Thr Arg Gly Tyr Lys Ser Leu Ala Gly
                 25                  30                  35

TGT CTT GGC CAT GGT CCC CTG GTG CTG CAA CTC CTC TCC TTC ACG CTC        197
Cys Leu Gly His Gly Pro Leu Val Leu Gln Leu Leu Ser Phe Thr Leu
             40                  45                  50

TTG GCT GGG CTC CTT GTC CAA GTG TCC AAG GTC CCC AGC TCC ATA AGT        245
Leu Ala Gly Leu Leu Val Gln Val Ser Lys Val Pro Ser Ser Ile Ser
         55                  60                  65

CAG GAA CAA TCC AGG CAA GAC GCG ATC TAC CAG AAC CTG ACC CAG CTT        293
Gln Glu Gln Ser Arg Gln Asp Ala Ile Tyr Gln Asn Leu Thr Gln Leu
     70                  75                  80

AAA GCT GCA GTG GGT GAG CTC TCA GAG AAA TCC AAG CTG CAG GAG ATC        341
Lys Ala Ala Val Gly Glu Leu Ser Glu Lys Ser Lys Leu Gln Glu Ile
 85                  90                  95                  100
```

```
TAC CAG GAG CTG ACC CAG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG      389
Tyr Gln Glu Leu Thr Gln Leu Lys Ala Ala Val Gly Glu Leu Pro Glu
                105                 110                 115

AAA TCT AAG CTG CAG GAG ATC TAC CAG GAG CTG ACC CGG CTG AAG GCT      437
Lys Ser Lys Leu Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala
            120                 125                 130

GCA GTG GGT GAG CTT CCA GAG AAA TCT AAG CTG CAG GAG ATC TAC CAG      485
Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Leu Gln Glu Ile Tyr Gln
        135                 140                 145

GAG CTG ACC TGG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG AAA TCT      533
Glu Leu Thr Trp Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser
    150                 155                 160

AAG ATG CAG GAG ATC TAC CAG GAG CTG ACT CGG CTG AAG GCT GCA GTG      581
Lys Met Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val
165                 170                 175                 180

GGT GAG CTT CCA GAG AAA TCT AAG CAG CAG GAG ATC TAC CAG GAG CTG      629
Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu
                185                 190                 195

ACC CGG CTG AAG GCT GCA GTG GGT GAG CTT CCA GAG AAA TCT AAG CAG      677
Thr Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln
            200                 205                 210

CAG GAG ATC TAC CAG GAG CTG ACC CGG CTG AAG GCT GCA GTG GGT GAG      725
Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val Gly Glu
        215                 220                 225

CTT CCA GAG AAA TCT AAG CAG CAG GAG ATC TAC CAG GAG CTG ACC CAG      773
Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu Thr Gln
    230                 235                 240

CTG AAG GCT GCA GTG GAA CGC CTG TGC CAC CCC TGT CCC TGG GAA TGG      821
Leu Lys Ala Ala Val Glu Arg Leu Cys His Pro Cys Pro Trp Glu Trp
245                 250                 255                 260

ACA TTC TTC CAA GGA AAC TGT TAC TTC ATG TCT AAC TCC CAG CGG AAC      869
Thr Phe Phe Gln Gly Asn Cys Tyr Phe Met Ser Asn Ser Gln Arg Asn
                265                 270                 275

TGG CAC GAC TCC ATC ACC GCC TGC AAA GAA GTG GGG GCC CAG CTC GTC      917
Trp His Asp Ser Ile Thr Ala Cys Lys Glu Val Gly Ala Gln Leu Val
            280                 285                 290

GTA ATC AAA AGT GCT GAG GAG CAG AAC TTC CTA CAG CTG CAG TCT TCC      965
Val Ile Lys Ser Ala Glu Glu Gln Asn Phe Leu Gln Leu Gln Ser Ser
        295                 300                 305
```

```
AGA AGT AAC CGC TTC ACC TGG ATG GGA CTT TCA GAT CTA AAT CAG GAA      1013
Arg Ser Asn Arg Phe Thr Trp Met Gly Leu Ser Asp Leu Asn Gln Glu
    310                 315                 320

GGC ACG TGG CAA TGG GTG GAC GGC TCA CCT CTG TTG CCC AGC TTC AAG      1061
Gly Thr Trp Gln Trp Val Asp Gly Ser Pro Leu Leu Pro Ser Phe Lys
325                 330                 335                 340

CAG TAT TGG AAC AGA GGA GAG CCC AAC AAC GTT GGG GAG GAA GAC TGC      1109
Gln Tyr Trp Asn Arg Gly Glu Pro Asn Asn Val Gly Glu Glu Asp Cys
                345                 350                 355

GCG GAA TTT AGT GGC AAT GGC TGG AAC GAC GAC AAA TGT AAT CTT GCC      1157
Ala Glu Phe Ser Gly Asn Gly Trp Asn Asp Asp Lys Cys Asn Leu Ala
            360                 365                 370

AAA TTC TGG ATC TGC AAA AAG TCC GCA GCC TCC TGC TCC AGG GAT GAA      1205
Lys Phe Trp Ile Cys Lys Lys Ser Ala Ala Ser Cys Ser Arg Asp Glu
        375                 380                 385

GAA CAG TTT CTT TCT CCA GCC CCT GCC ACC CCA AAC CCC CCT CCT GCG      1253
Glu Gln Phe Leu Ser Pro Ala Pro Ala Thr Pro Asn Pro Pro Pro Ala
    390                 395                 400

TAGCAGAACT TCACCCCCTT TTAAGCTACA GTTCCTTCTC TCCATCCTTC GACCTTTAG     1312
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 404 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ser Asp Ser Lys Glu Pro Arg Leu Gln Gln Leu Gly Leu Leu Glu
  1               5                  10                  15

Glu Glu Gln Leu Arg Gly Leu Gly Phe Arg Gln Thr Arg Gly Tyr Lys
             20                  25                  30

Ser Leu Ala Gly Cys Leu Gly His Gly Pro Leu Val Leu Gln Leu Leu
         35                  40                  45

Ser Phe Thr Leu Leu Ala Gly Leu Leu Val Gln Val Ser Lys Val Pro
     50                  55                  60
```

```
Ser Ser Ile Ser Gln Glu Gln Ser Arg Gln Asp Ala Ile Tyr Gln Asn
 65                  70                  75                  80

Leu Thr Gln Leu Lys Ala Ala Val Gly Glu Leu Ser Glu Lys Ser Lys
                 85                  90                  95

Leu Gln Glu Ile Tyr Gln Glu Leu Thr Gln Leu Lys Ala Ala Val Gly
            100                 105                 110

Glu Leu Pro Glu Lys Ser Lys Leu Gln Glu Ile Tyr Gln Glu Leu Thr
        115                 120                 125

Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Leu Gln
    130                 135                 140

Glu Ile Tyr Gln Glu Leu Thr Trp Leu Lys Ala Ala Val Gly Glu Leu
145                 150                 155                 160

Pro Glu Lys Ser Lys Met Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu
                165                 170                 175

Lys Ala Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile
            180                 185                 190

Tyr Gln Glu Leu Thr Arg Leu Lys Ala Ala Val Gly Glu Leu Pro Glu
        195                 200                 205

Lys Ser Lys Gln Gln Glu Ile Tyr Gln Glu Leu Thr Arg Leu Lys Ala
    210                 215                 220

Ala Val Gly Glu Leu Pro Glu Lys Ser Lys Gln Gln Glu Ile Tyr Gln
225                 230                 235                 240

Glu Leu Thr Gln Leu Lys Ala Ala Val Glu Arg Leu Cys His Pro Cys
                245                 250                 255

Pro Trp Glu Trp Thr Phe Phe Gln Gly Asn Cys Tyr Phe Met Ser Asn
            260                 265                 270

Ser Gln Arg Asn Trp His Asp Ser Ile Thr Ala Cys Lys Glu Val Gly
        275                 280                 285

Ala Gln Leu Val Val Ile Lys Ser Ala Glu Glu Gln Asn Phe Leu Gln
    290                 295                 300

Leu Gln Ser Ser Arg Ser Asn Arg Phe Thr Trp Met Gly Leu Ser Asp
305                 310                 315                 320

Leu Asn Gln Glu Gly Thr Trp Gln Trp Val Asp Gly Ser Pro Leu Leu
                325                 330                 335

Pro Ser Phe Lys Gln Tyr Trp Asn Arg Gly Glu Pro Asn Asn Val Gly
            340                 345                 350
```

```
Glu Glu Asp Cys Ala Glu Phe Ser Gly Asn Gly Trp Asn Asp Asp Lys
        355                 360                 365

Cys Asn Leu Ala Lys Phe Trp Ile Cys Lys Lys Ser Ala Ala Ser Cys
    370                 375                 380

Ser Arg Asp Glu Glu Gln Phe Leu Ser Pro Ala Pro Ala Thr Pro Asn
385                 390                 395                 400

Pro Pro Pro Ala
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 127 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Human immunodeficiency virus type 1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Cys His Pro Cys Pro Trp Glu Trp Thr Phe Phe Gln Gly Asn Cys Tyr
1               5                   10                  15

Phe Met Ser Asn Ser Gln Arg Asn Trp His Asp Ser Ile Thr Ala Cys
            20                  25                  30

Lys Glu Val Gly Ala Gln Leu Val Val Ile Lys Ser Ala Glu Glu Gln
        35                  40                  45

Asn Phe Leu Gln Leu Gln Ser Ser Arg Ser Asn Arg Phe Thr Trp Met
    50                  55                  60

Gly Leu Ser Asp Leu Asn Gln Glu Gly Thr Trp Gln Trp Val Asp Gly
65                  70                  75                  80

Ser Pro Leu Leu Pro Ser Phe Lys Gln Tyr Trp Asn Arg Gly Glu Pro
                85                  90                  95
```

```
Asn Asn Val Gly Glu Glu Asp Cys Ala Glu Phe Ser Gly Asn Gly Trp
            100                 105                 110

Asn Asp Asp Lys Cys Asn Leu Ala Lys Phe Trp Ile Cys Lys Lys
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 126 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Cys Gly Ala Gln Ser Arg Gln Trp Glu Tyr Phe Glu Gly Arg Cys Tyr
1               5                   10                  15

Tyr Phe Ser Leu Ser Arg Met Ser Trp His Lys Ala Lys Ala Glu Cys
            20                  25                  30

Glu Glu Met His Ser His Leu Ile Ile Ile Asp Ser Tyr Ala Lys Gln
        35                  40                  45

Asn Phe Val Met Phe Arg Thr Arg Asn Glu Arg Phe Trp Ile Gly Leu
    50                  55                  60

Thr Asp Glu Asn Gln Glu Gly Glu Trp Gln Trp Val Asp Gly Thr Asp
65                  70                  75                  80

Thr Arg Ser Ser Phe Thr Phe Trp Lys Glu Gly Glu Pro Asn Asn Arg
                85                  90                  95

Gly Phe Asn Glu Asp Cys Ala His Val Trp Thr Ser Gly Gln Trp Asn
            100                 105                 110

Asp Val Tyr Cys Thr Tyr Glu Cys Tyr Tyr Val Cys Glu Lys
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Cys Asn Thr Cys Pro Glu Lys Trp Ile Asn Phe Gln Arg Lys Cys Tyr
1               5                  10                  15

Tyr Phe Gly Lys Gly Thr Lys Gln Trp Val His Ala Arg Tyr Ala Cys
            20                  25                  30

Asp Asp Met Glu Gly Gln Leu Val Ser Ile His Ser Pro Glu Glu Gln
        35                  40                  45

Asp Phe Leu Thr Lys His Ala Ser His Thr Gly Ser Trp Ile Gly Leu
    50                  55                  60

Arg Asn Leu Asp Leu Lys Gly Glu Phe Ile Trp Val Asp Gly Ser His
65                  70                  75                  80

Val Asp Tyr Ser Asn Trp Ala Pro Gly Glu Pro Thr Ser Arg Ser Gln
                85                  90                  95

Gly Glu Asp Cys Val Met Met Arg Gly Ser Gly Arg Trp Asn Asp Ala
            100                 105                 110

Phe Cys Asp Arg Lys Leu Gly Ala Trp Val Cys Asp Arg
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Arg Thr Cys Cys Pro Val Asn Trp Val Glu His Glu Arg Ser Cys Tyr
1               5                  10                  15

Trp Phe Ser Arg Ser Gly Lys Ala Trp Ala Asp Ala Asp Asn Tyr Cys
            20                  25                  30

Arg Leu Glu Asp Ala His Leu Val Val Val Thr Ser Trp Glu Glu Gln
        35                  40                  45

Lys Phe Val Gln His His Ile Gly Pro Val Asn Thr Trp Met Gly Leu
    50                  55                  60

His Asp Gln Asn Gly Pro Trp Lys Trp Val Asp Gly Thr Asp Tyr Glu
65                  70                  75                  80

Thr Gly Phe Lys Asn Trp Arg Pro Glu Gln Pro Asp Asp Trp Tyr Gly
                85                  90                  95

His Gly Leu Gly Gly Gly Glu Asp Cys Ala His Phe Thr Asp Asp Gly
            100                 105                 110

Arg Trp Asn Asp Asp Val Cys Gln Arg Pro Tyr Arg Trp Val Cys Glu
        115                 120                 125

Thr
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Arg Thr Cys Cys Pro Val Asn Trp Val Glu His Gln Gly Ser Cys Tyr
1               5                  10                  15

Trp Phe Ser His Ser Gly Lys Ala Trp Ala Glu Ala Glu Lys Tyr Cys
            20                  25                  30

Gln Leu Glu Asn Ala His Leu Val Val Ile Asn Ser Trp Glu Glu Gln
        35                  40                  45

Lys Phe Ile Val Gln His Thr Asn Pro Phe Asn Thr Trp Ile Gly Leu
    50                  55                  60

Thr Asp Ser Asp Gly Ser Trp Lys Trp Val Asp Gly Thr Asp Tyr Arg
65                  70                  75                  80

His Asn Tyr Lys Asn Trp Ala Val Thr Gln Pro Asp Asn Trp His Gly
                85                  90                  95

His Glu Leu Gly Gly Ser Glu Asp Cys Val Glu Val Gln Pro Asp Gly
            100                 105                 110

Arg Trp Asn Asp Asp Phe Cys Leu Gln Val Tyr Arg Trp Val Cys Glu
        115                 120                 125

Lys
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Leu Gln Leu Ile Met Gln Asp Trp Lys Tyr Phe Asn Gly Lys Phe Tyr
1               5                  10                  15
```

```
Tyr Phe Ser Arg Asp Lys Lys Ser Trp His Glu Ala Glu Asn Phe Cys
            20                  25                  30

Val Ser Gln Gly Ala His Leu Ala Ser Val Thr Ser Gln Glu Glu Gln
        35                  40                  45

Ala Phe Leu Val Gln Ile Thr Asn Ala Val Asp His Trp Ile Gly Leu
    50                  55                  60

Thr Asp Gln Gly Thr Glu Gly Asn Trp Arg Trp Val Asp Gly Thr Pro
65                  70                  75                  80

Phe Asp Tyr Val Gln Ser Arg Arg Phe Trp Arg Lys Gly Gln Pro Asp
                85                  90                  95

Asn Trp Arg His Gly Asn Gly Glu Arg Glu Asp Cys Val His Leu Gln
            100                 105                 110

Arg Met Trp Asn Asp Met Ala Cys Gly Thr Ala Tyr Asn Trp Val Cys
        115                 120                 125

Lys Lys
    130
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
Pro Thr His Cys Pro Ser Gln Trp Trp Pro Tyr Ala Gly His Cys Tyr
1               5                   10                  15

Lys Ile His Arg Asp Glu Lys Lys Ile Gln Arg Asp Ala Leu Thr Thr
            20                  25                  30

Cys Arg Lys Glu Gly Gly Asp Leu Thr Ser Ile His Thr Ile Glu Glu
        35                  40                  45
```

Leu Asp Phe Ile Ile Ser Gln Leu Gly Leu Glu Pro Asn Asp Glu Leu
50 55 60

Trp Ile Gly Leu Asn Asp Ile Lys Ile Gln Met Tyr Phe Glu Trp Ser
65 70 75 80

Asp Gly Thr Pro Val Thr Phe Thr Lys Trp Leu Arg Gly Glu Pro Ser
85 90 95

His Glu Asn Asn Arg Gln Glu Asp Cys Val Val Met Lys Gly Lys Asp
100 105 110

Gly Tyr Trp Ala Asp Arg Gly Cys Glu Trp Pro Leu Gly Tyr Ile Cys
115 120 125

Lys Met
130

We claim:

1. A method of inhibiting HIV infection of mammalian cells comprising contacting the cells with an effective amount of a compound selected from the group consisting of a mannose carbohydrate, a fucose carbohydrate, a lectin and a drug, for a time period sufficient to significantly inhibit the binding of HIV to a non-CD4 cell surface protein.

2. The method of Claim 1, wherein the non-CD4 cell surface protein is a gp120 receptor having a specific binding affinity for gp120 of about Kd = 1.3 nM to about Kd = 2.0 nM.

3. The method of Claim 2, wherein the gp120 receptor is present on placental cells.

4. The method of Claim 2, wherein the gp120 receptor is present on muscle cells.

5. The method of Claim 2, wherein the gp120 receptor is present on neural cells.

6. The method of Claim 5, wherein the neural cells are brain cells.

7. The method of Claim 5, wherein the neural cells are dendritic cells.

8. The method of Claim 2, wherein the gp120 receptor is present on mucosal cells.

10. The method of Claim 1, wherein the compound is mannose.

10. The method of Claim 1, wherein the compound is fucose.

11. The method of Claim 1, wherein the compound is a mannose-containing carbohydrate.

12. The method of Claim 11, where the carbohydrate is mannan.

13. The method of Claim 1, wherein the compound is a pradimicin A antibiotic.

14. A substantially purified non-CD4 gp120 receptor protein comprising a protein substantially corresponding to a non-CD4 mammalian cell surface protein that has a

specific binding affinity for gp120, said protein containing about 400 amino acid residues, having a molecular weight of about 45,000 daltons and having a binding affinity for gp120 characterized by a Kd of about 1.3 nM to about 2 nM.

15. The gp120 receptor protein of Claim 14, wherein the binding of the gp120 receptor protein to gp120 is inhibited by a compound selected from the group consisting of a mannose carbohydrate, a fucose carbohydrate, a lectin and a drug.

16. The gp120 receptor of Claim 15, wherein the compound is mannose.

17. The gp120 receptor protein of Claim 15, wherein the compound is a pradimicin A antibiotic.

18. The gp120 receptor protein of Claim 14, wherein the protein is produced by recombinant means.

19. The gp120 receptor protein of Claim 18, wherein said recombinant means comprises the cloning of a cDNA isolated from a library of recombinant placental genes.

20. A DNA molecule encoding the gp120 receptor protein of Claim 14, wherein the DNA is a complementary DNA that transcribes an mRNA found in cells selected from the group consisting of placental cells, brain cells, muscle cells and colon cells.

21. A method of detecting the presence of HIV in a sample comprising:

(a) admixing in an aqueous medium a sample to be assayed with a non-CD4 gp120 receptor protein having a specific binding affinity for gp120 characterized by a Kd of about 1.3 nM to about 2.0 nM in an amount sufficient to carry out at least one assay;

(b) maintaining the admixture for a time period sufficient for the gp120 receptor protein to bind to any HIV present in the sample and form a reaction product; and

(c) determining the presence of the HIV containing reaction product.

22. The method of Claim 21, wherein the gp120 receptor protein contains about 400 amino acid residues and has a molecular weight of about 45,000 daltons.

23. The method of Claim 21, wherein the gp120 receptor protein is affixed to a solid matrix to form a solid support.

24. The method of Claim 21, wherein the presence of the reaction product is determined by contacting the sample with a reagent capable of detecting the bound gp120 receptor protein.

25. The method of Claim 24, wherein the reagent is a labelled antibody directed against the gp120 receptor protein.

26. A diagnostic system in kit form, for assaying for the presence of HIV in a fluid sample, comprising a package containing a non-CD4 receptor protein having a specific affinity for gp120 characterized by a Kd of about 1.3 nM to about 2.0 nM, and instructions for use.

27. The diagnostic system of Claim 26, wherein the non-CD4 gp120 receptor protein is affixed to a solid matrix to form a solid support.

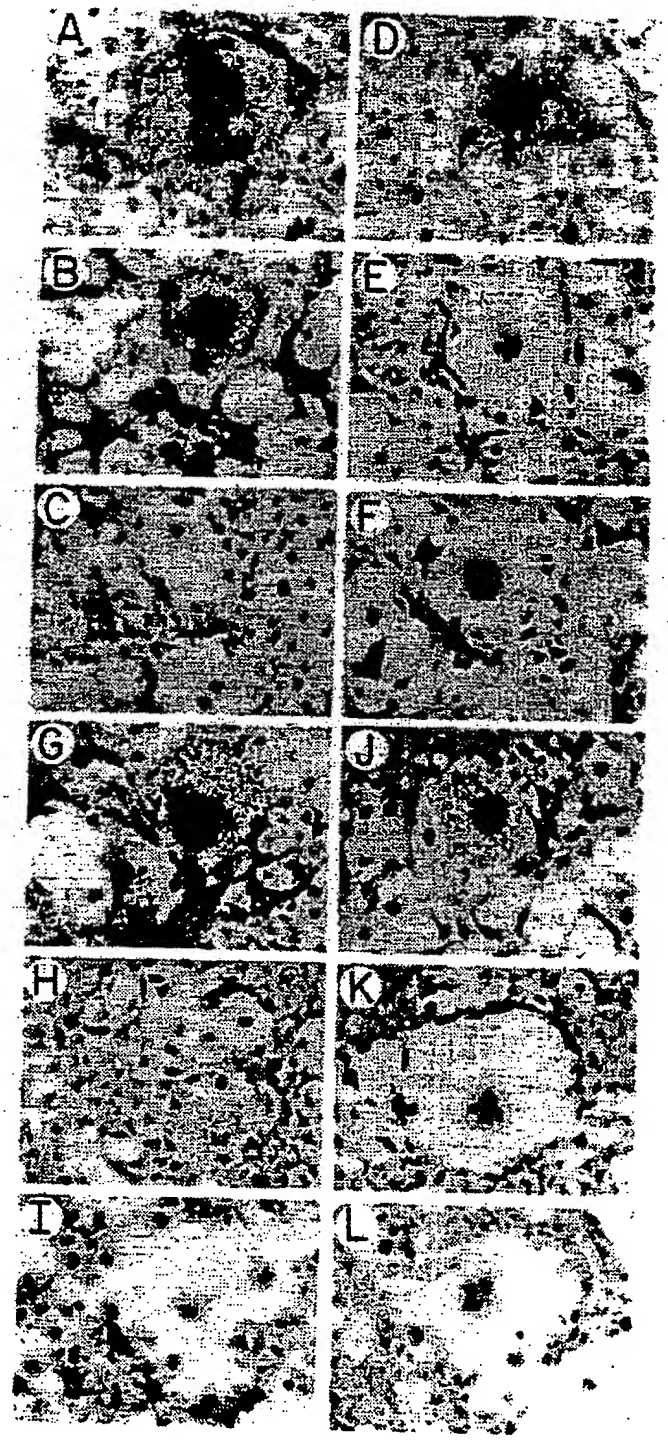


Figure 1A

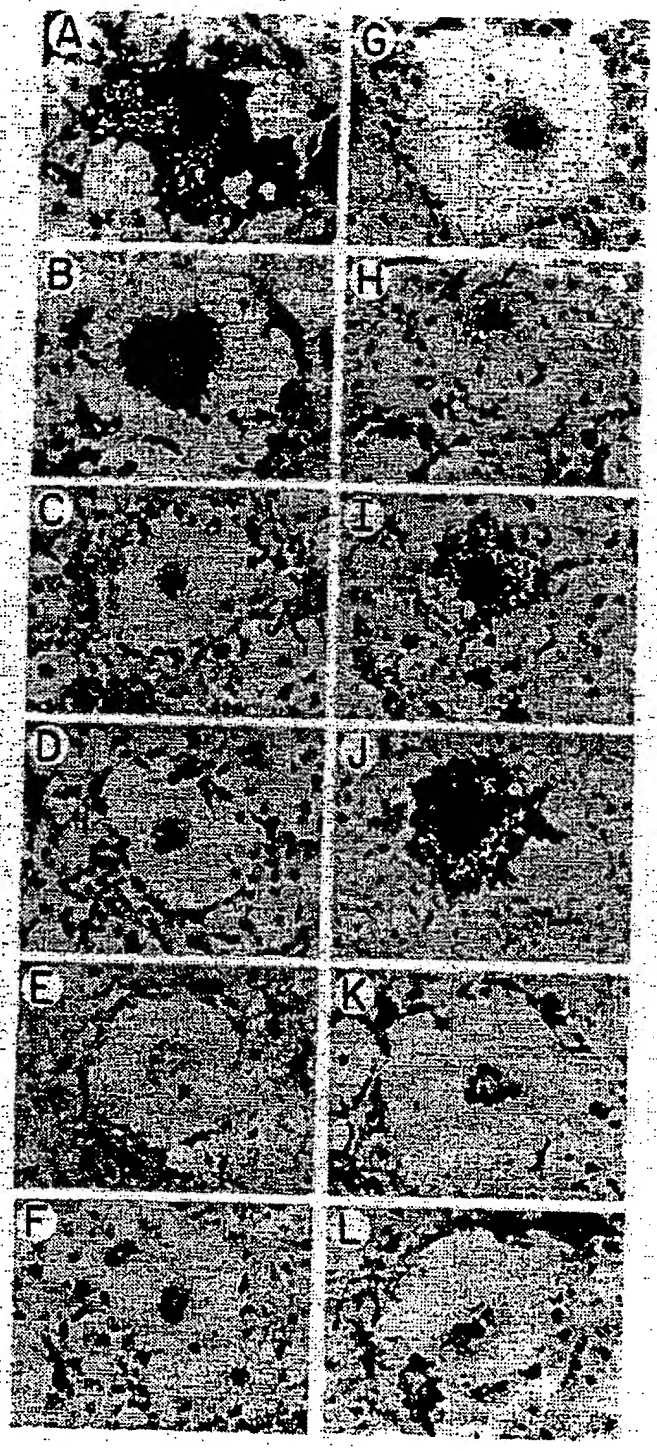


Figure 1B

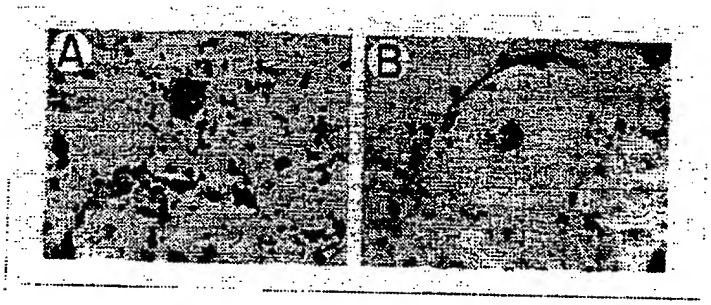


Figure 1C

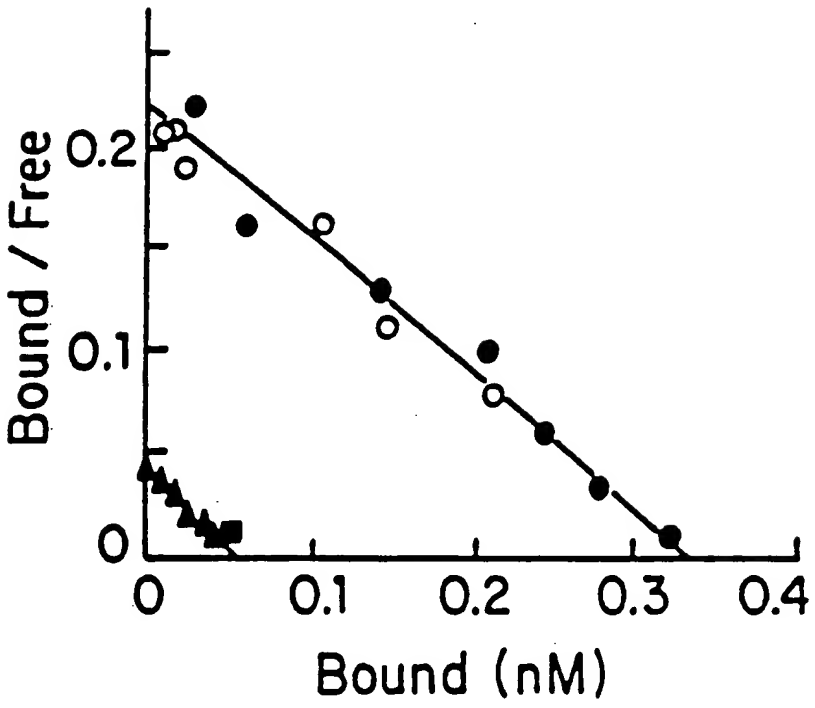


Figure 2A

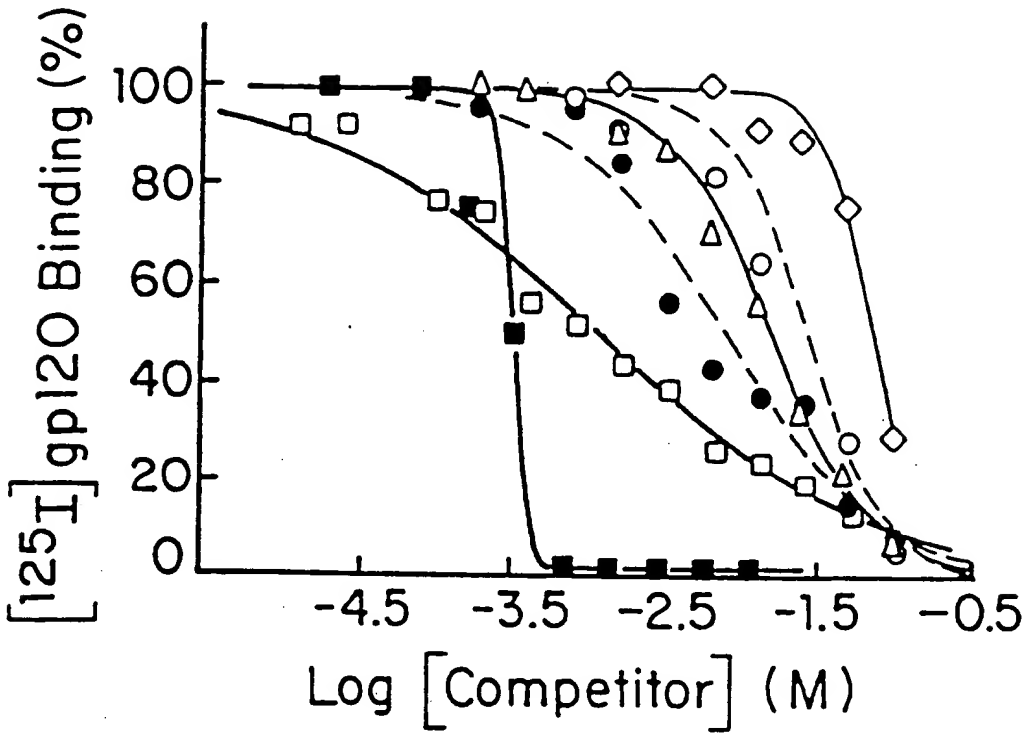

[125I]gp120 Binding (%)
100
80
60
40
20
0
-4.5
-3.5
-2.5
-1.5
-0.5
Log [Competitor] (M)

Figure 2B

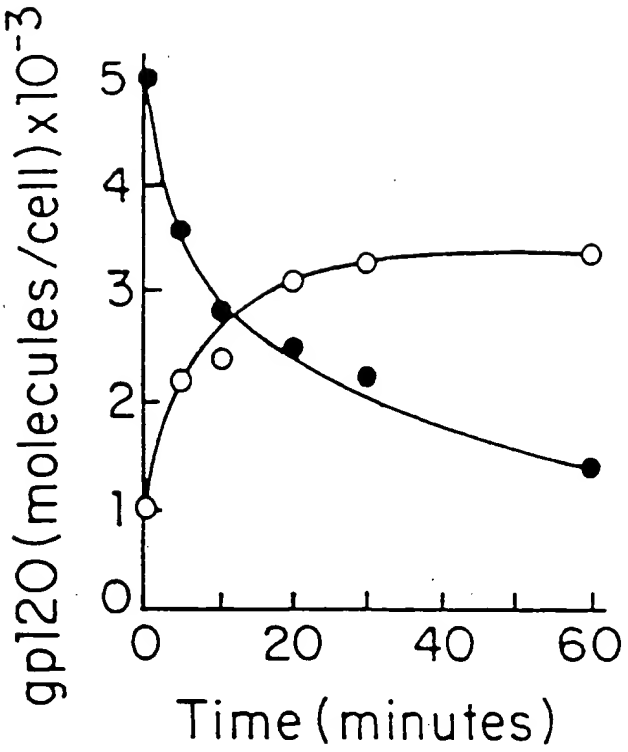


Figure 2C

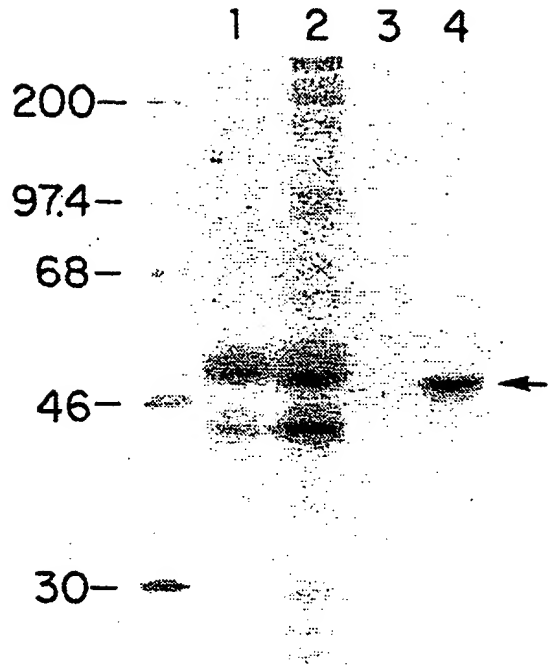


Figure 2D

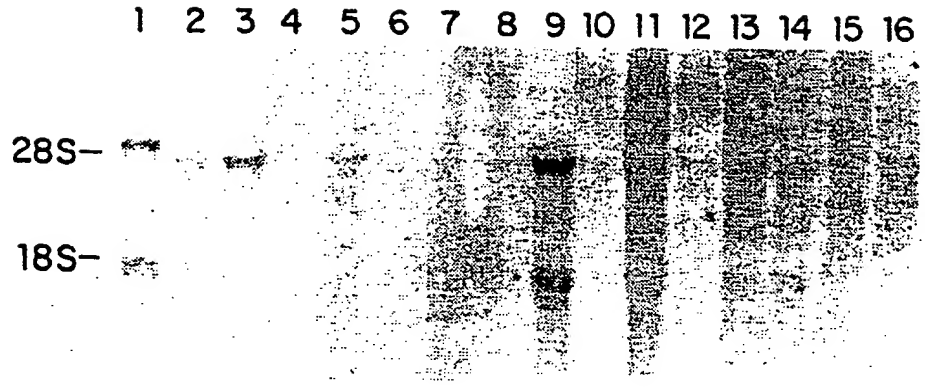


Figure 2E

```
   1                                                CTAAAGCAGGAGTTCTGGACACTGGGGGAGAGTGGGGTGAC

  42   ATGAGTGACTCCAAGGAACCAAGACTGCAGCAGCTGGGCCTCCTGGAGGAGGAACAGCTG
   1   M  S  D  S  K  E  P  R  L  Q  Q  L  G  L  L  E  E  E  Q  L

 102   AGAGGCCTTGGATTCCGACAGACTCGAGGATACAAGAGCTTAGCAGGGTGTCTTGGCCAT
  21   R  G  L  G  F  R  Q  T  R  G  Y  K  S  L  A  G  C  L  G  H

 162   GGTCCCCTGGTGCTGCAACTCCTCTCCTTCACGCTCTTGGCTGGGCTCCTTGTCCAAGTG
  41   G  P  L  V  L  Q  L  L  S  F  T  L  L  A  G  L  L  V  Q  V

 222   TCCAAGGTCCCCAGCTCCATAAGTCAGGAACAATCCAGGCAAGACGCGATCTACCAGAAC
  61   S  K  V  P  S  S  I  S  Q  E  Q  S  R  Q  D  A  I  Y  Q  N

                                                          R1       *
 282   CTGACCCAGCTTAAAGCTGCAGTGGGTGAGCTCTCAGAGAAATCCAAGCTGCAGGAGATC
  81   L  T  Q  L  K  A  A  V  G  E  L  S  E  K  S  K  L  Q  E  I

                                                                R2
 342   TACCAGGAGCTGACCCAGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCTG
 101   Y  Q  E  L  T  Q  L  K  A  A  V  G  E  L  P  E  K  S  K  L

 402   CAGGAGATCTACCAGGAGCTGACCCGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAA
 121   Q  E  I  Y  Q  E  L  T  R  L  K  A  A  V  G  E  L  P  E  K

           R3
 462   TCTAAGCTGCAGGAGATCTACCAGGAGCTGACCTGGCTGAAGGCTGCAGTGGGTGAGCTT
 141   S  K  L  Q  E  I  Y  Q  E  L  T  W  L  K  A  A  V  G  E  L

                     R4
 522   CCAGAGAAATCTAAGATGCAGGAGATCTACCAGGAGCTGACTCGGCTGAAGGCTGCAGTG
 161   P  E  K  S  K  M  Q  E  I  Y  Q  E  L  T  R  L  K  A  A  V

                                    R5
 582   GGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAGGAGCTGACCCGGCTGAAG
 181   G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q  E  L  T  R  L  K

                                          R6
 642   GCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAGGAGCTGACC
 201   A  A  V  G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q  E  L  T

                                                   R7
 702   CGGCTGAAGGCTGCAGTGGGTGAGCTTCCAGAGAAATCTAAGCAGCAGGAGATCTACCAG
 221   R  L  K  A  A  V  G  E  L  P  E  K  S  K  Q  Q  E  I  Y  Q

                                                               R8
 762   GAGCTGACCCAGCTGAAGGCTGCAGTGGAACGCCTGTGCCACCCCTGTCCCTGGGAATGG
 241   E  L  T  Q  L  K  A  A  V  E  R  L  C  H  P  C  P  W  E  W
                                                L
 822   ACATTCTTCCAAGGAAACTGTTACTTCATGTCTAACTCCCAGCGGAACTGGCACGACTCC
 261   T  F  F  Q  G  N  C  Y  F  M  S  N  S  Q  R  N  W  H  D  S
```

Figure 3A

```
882   ATCACCGCCTGCAAAGAAGTGGGGGCCCAGCTCGTCGTAATCAAAAGTGCTGAGGAGCAG
281   I  T  A  C  K  E  V  G  A  Q  L  V  V  I  K  S  A  E  E  Q

942   AACTTCCTACAGCTGCAGTCTTCCAGAAGTAACCGCTTCACCTGGATGGGACTTTCAGAT
301   N  F  L  Q  L  Q  S  S  R  S  N  R  F  T  W  M  G  L  S  D

1002  CTAAATCAGGAAGGCACGTGGCAATGGGTGGACGGCTCACCTCTGTTGCCCAGCTTCAAG
321   L  N  Q  E  G  T  W  Q  W  V  D  G  S  P  L  L  P  S  F  K

1062  CAGTATTGGAACAGAGGAGAGCCCAACAACGTTGGGGAGGAAGACTGCGCGGAATTTAGT
341   Q  Y  W  N  R  G  E  P  N  N  V  G  E  E  D  C  A  E  F  S

1122  GGCAATGGCTGGAACGACGACAAATGTAATCTTGCCAAATTCTGGATCTGCAAAAAGTCC
361   G  N  G  W  N  D  D  K  C  N  L  A  K  F  W  I  C  K  K  S

1182  GCAGCCTCCTGCTCCAGGGATGAAGAACAGTTTCTTTCTCCAGCCCCTGCCACCCCAAAC
381   A  A  S  C  S  R  D  E  E  Q  F  L  S  P  A  P  A  T  P  N

1242  CCCCCTCCTGCGTAGCAGAACTTCACCCCCTTTTAAGCTACAGTTCCTTCTCTCCATCCT
401   P  P  P  A  ***

1302  TCGACCTTTAG
```

Figure 3A (cont.)

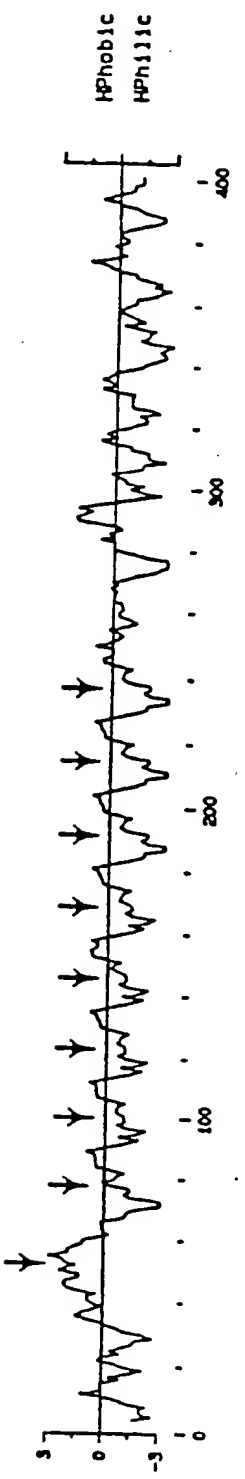


Figure 3B

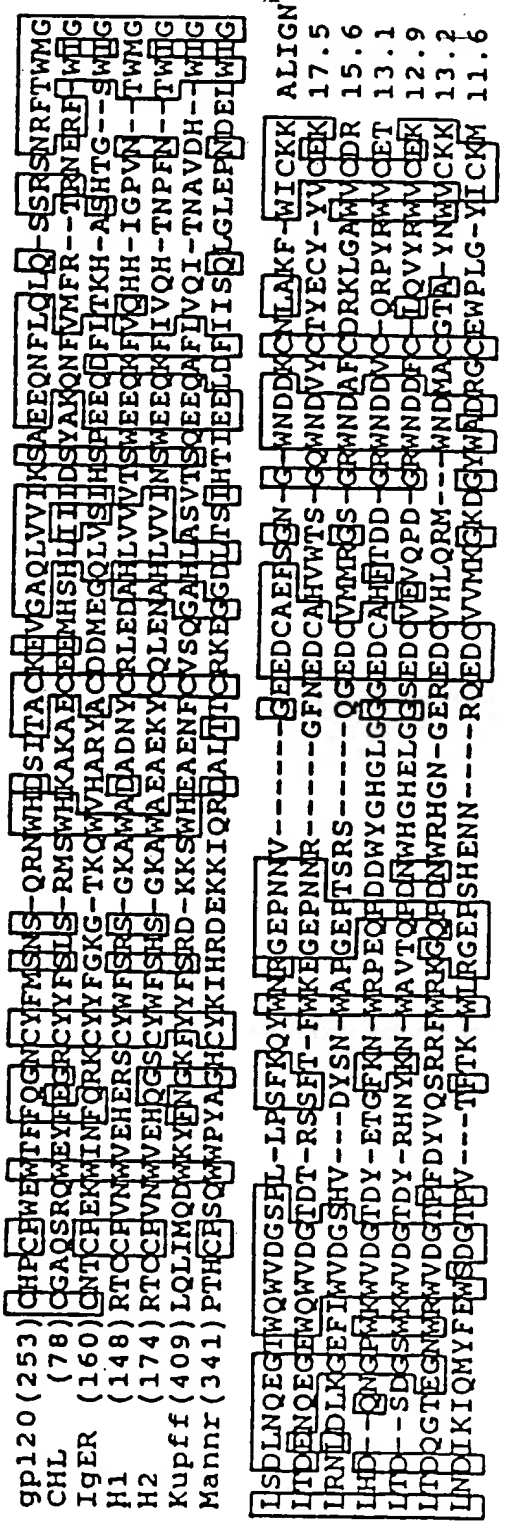


Figure 3C

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8356–8360, September 1992
Medical Sciences

Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120

(virus receptor/type II membrane protein/mannose binding)

BENSON M. CURTIS*, SONYA SCHARNOWSKE, AND ANDREW J. WATSON†

Bristol-Myers Squibb Pharmaceutical Research Institute, 3005 First Avenue, Seattle, WA 98121

Communicated by George J. Todaro, May 26, 1992

ABSTRACT **The binding of the human immunodeficiency virus (HIV) envelope glycoprotein gp120 to the cell surface receptor CD4 has been considered a primary determinant of viral tropism. A number of cell types, however, can be infected by the virus, or bind gp120, in the absence of CD4 expression. Human placenta was identified as a tissue that binds gp120 in a CD4-independent manner. A placental cDNA library was screened by expression cloning and a cDNA (clone 11) encoding a gp120-binding protein unrelated to CD4 was isolated. The 1.3-kilobase cDNA predicts a protein of 404 amino acids with a calculated M_r of 45,775 and organized into three domains: an N-terminal cytoplasmic and hydrophobic region, a set of seven complete and one incomplete tandem repeat, and a C-terminal domain with homology to C-type (calcium-dependent) lectins. A type II membrane orientation (N-terminal cytoplasmic) is predicted both by the cDNA sequence and by the reactivity of C-terminal peptide-specific antiserum with the surface of clone 11 transfected cells. Native and recombinant gp120 and whole virus bind transfected cells. gp120 binding is high affinity (K_d, 1.3–1.6 nM) and inhibited by mannan, D-mannose, and L-fucose; once bound, gp120 is internalized rapidly. Collectively, these data demonstrate that the gp120-binding protein is a membrane-associated mannose-binding lectin. Proteins of this type may play an important role in the CD4-independent association of HIV with cells.**

One of the first steps in the infection of T cells with human immunodeficiency virus (HIV) is binding of the envelope glycoprotein gp120 to the differentiation antigen CD4 (see ref. 1 for review). The observation of HIV infection of (2–7), and gp120 binding to (8), a number of cell types in the absence of detectable CD4 expression suggests that CD4-independent mechanisms of viral entry also exist. This apparent absence of a strict requirement for CD4 potentially broadens the tissue tropism of the virus. In addition, direct infection by HIV may not always be required to elicit cytopathic effects. For example, CD4-independent binding can occur in neural tissue (8, 9), and exposure of neuronal cultures to gp120 can result in cytotoxicity (9, 10).

The identification of non-CD4 HIV receptors is important if the diverse clinical manifestations observed in HIV infection are to be understood. In this report we describe the use of a eukaryotic expression system (11, 12) to screen cDNAs derived from human placenta, a tissue that exhibits CD4-independent binding of gp120. A cDNA clone was isolated that encodes a gp120-binding protein distinct from CD4. This protein has structural features and binding characteristics that indicate it is a member of the family of C-type mannose-binding proteins.‡

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Expression Cloning. Pools of 90,000 cDNAs from a placental pCDM8 library (a gift from B. Seed, Harvard Medical School) were transfected by electroporation into COS-7 cells. After 3 days, transfected cells were screened for binding with 1 nM ^{125}I-labeled recombinant vaccinia virus-derived gp120 (vgp120) (refs. 8, 11–13; A. Blomstedt, S. Olofsson, E. Sjogren-Jansson, S. Jeansson, L. Akerblom, J.-E. S. Hansen, and S.-L. Hu, personal communication) after a 1-hr preincubation with CD4a antibody G17-2 (5 μg/ml) by visual inspection of single cells after autoradiography (3-day exposure). [Antibody G17-2 belongs to the CD4a subgroup of CD4 antibodies that block both gp120 binding to CD4 and viral infection (15).] After ≈30 pools had been screened a positive pool was identified and rescreened as successively smaller pools to yield a single cDNA (clone 11). Cells expressing CD4 were obtained following transfection with an equal amount of πH3MCD4 (16). Specificity to gp120 binding was assigned by binding of gp120 purified from HIV_{BRU} (native gp120, ngp120) (17), block of binding by baculovirus-derived gp120 (bgp120) (American Biotechnologies, Columbia, MD), and elimination of binding by immunoprecipitation of the ^{125}I-labeled gp120 preparation with the anti-gp120 monoclonal antibody 110.1 (15), anti-mouse IgG, and protein A-Sepharose. Untransfected COS cells did not display a density of silver grains greater than the background.

Sequencing and Analysis. Clone 11 cDNA in pCDM8 was sequenced on both strands by the dideoxy chain-termination method. Hydropathy was assigned by a Kyte–Doolittle plot (7-residue window) obtained with the Wisconsin Genetics Computer Group package, and sequence alignments and ALIGN scores were generated using PC/GENE.

Ligand Binding, Inhibition, and Internalization Assays. Binding assays were conducted essentially as described (12). For inhibition assays, transfected COS cells or gp120 was preincubated for 1 hr with inhibitor. In ligand internalization assays, transfected COS cells were incubated with 1 nM ^{125}I-gp120 for 5 hr at 4°C, washed, and incubated at 37°C for the time indicated; surface and internalized gp120 were separated by acid treatment (18).

Stable Transfection of HeLa Cells. Clone 11 cDNA was inserted in the *Hin*dIII/*Not* I sites of pcDNAI/Neo vector (Invitrogen, San Diego). HeLa cells were transfected by a calcium phosphate procedure and, after 3 days, selected with Geneticin (GIBCO). Resistant cells were initially enriched for

Abbreviations: HIV, human immunodeficiency virus; vgp120, vaccinia virus-derived gp120; bgp120, baculovirus-derived gp120; ngp120, native gp120.

*Present address: Department of Experimental Medicine, University of Otago Medical School, Dunedin, New Zealand.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M98457).

expression of the gp120-binding protein by two rounds of sterile sorting on a Coulter flow cytometer following staining with fluorescein-labeled bgp120 (American Biotechnologies). Subsequent selection used vgp120 labeled by incubation with the anti-gp120 monoclonal antibody 110-4 (15), followed by a fluorescein-labeled anti-mouse Ig reagent. In total, four cycles of selection were used. Expression of gp120-binding protein over time was followed by staining with vgp120 (100 nM), antibody 110-4, and fluorescein-conjugated anti-mouse reagents. vgp120 binding was inhibited by preincubating cells with mannan (2–4 mg/ml; Sigma) for 30 min.

Generation of Rabbit Antisera. A peptide (564A) corresponding to the C terminus of the polypeptide encoded by clone 11 cDNA (Cys^{384}–Ala^{404}) was conjugated to ovalbumin and used to immunize rabbits. Sera were used following a second booster injection, and titers in peptide ELISAs exceeded 500,000.

RESULTS

cDNA Library Screening. Human placental membranes were found to bind recombinant vgp120 in the presence of antibodies that efficiently block gp120 association with CD4. Fifty to 90% of placental gp120 binding was estimated to be non-CD4. To attempt to identify the protein responsible, a placental cDNA library in the vector pCDM8 was screened by expression cloning procedures. COS cells were transfected with pools of cDNAs and CD4-independent gp120 binding activity was detected with radiolabeled vgp120 in the presence of the CD4a antibody G17-2, which blocks binding of gp120 to CD4. After ≈30 pools had been screened, a positive pool was identified and rescreened as successively smaller pools to yield a single cDNA (clone 11).

Affinity and Characteristics of gp120 Binding. Scatchard plots of gp120 binding to COS cells transfected with clone 11 cDNA gave a K_d of 1.7 ± 0.4 nM (n = 4) for vgp120 and 1.8 ± 0.2 nM (n = 4) for ngp120 (Fig. 1*A*), similar to the results obtained with isolated placental membranes (K_d = 1.3 nM) in the presence of CD4a antibodies (Fig. 1*A*). Calculations from the association and dissociation rate constants gave a similar comparative result. Concurrent analysis of gp120 binding to CD4 expressed on COS cells gave a K_d of 4–5 nM in agreement with previous reports (15, 19). Binding of vgp120 to clone 11 transfected cells was inhibited by bgp120 and ngp120 isolated from purified HIV_{BRU}. Undisrupted psoralen/UV-inactivated HIV_{BRU} also bound clone 11 transfected cells in a gp120-dependent manner (data not shown).

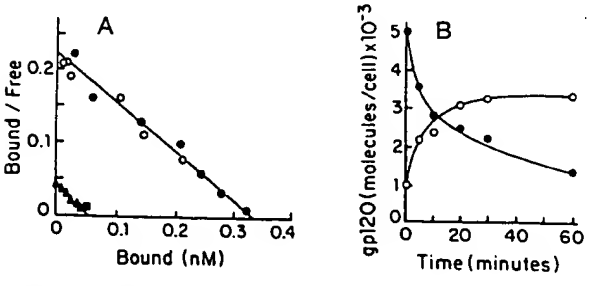


FIG. 1. Characterization of gp120 binding. (*A*) Scatchard analysis of ^{125}I-gp120 binding. ▲, vgp120 binding to placenta (K_d, 1.3 nM; B_{max}, 19 fmol/mg of protein); ■, with CD4a antibody (5 μg/ml); ○, vgp120 binding to clone 11 COS cells (K_d, 1.5 nM; B_{max}, 150,000 receptors per cell); ●, ngp120 (K_d, 1.6 nM; 149,000 receptors per cell). (*B*) Internalization of gp120 by clone 11-expressing COS cells. Points represent the mean of two experiments with vgp120 and ngp120. Transfected COS cells were incubated with ^{125}I-gp120 for 5 hr at 4°C prior to acid stripping procedures, which were conducted at 37°C. ●, Surface; ○, internal.

The fate of gp120 bound to the surface of transfected COS cells was also examined. In these experiments, ^{125}I-gp120 was incubated with cells for 5 hr at 4°C and internalization of the bound gp120 at 37°C was determined using acid stripping procedures (18) to remove cell-surface ^{125}I-gp120. The gp120 was rapidly converted to an acid-resistant form at 37°C (Fig. 1*B*), consistent with the ability of the gp120-binding protein to mediate ligand internalization into the cell.

Predicted Structure of gp120-Binding Protein. The 1.3-kilobase clone 11 cDNA encodes a protein of 404 amino acids with a calculated M_r of 45,775 (Fig. 2). No signal sequence is apparent, but a 21-residue hydrophobic tract (Gly^{41}–Ser^{61}) is present 40 residues from the N terminus (Fig. 3*A*). These features suggest a type II membrane orientation (N-terminal cytoplasmic), which is also supported by the distribution of positively charged amino acids within 15 residues of the hydrophobic region ["positive-inside" rule (20)]. A series of seven complete and one incomplete tandem repeat (Ile^{77}–Val^{249}) of nearly identical sequence follows. The remaining sequence, Cys^{253}–Ala^{404}, shows homology to C-type lectins (Fig. 3*B*): chick hepatic lectin (21), low-affinity IgE receptor (22), the asialoglycoprotein receptors [human H1 and H2 (23) are shown], the rat Kupffer cell receptor (24), and the human macrophage mannose receptor (25, 26).

Binding Inhibition Studies. The sequence homology of the gp120-binding protein to C-type lectins prompted evaluation of the role of sugars in recognition of gp120. Inhibition by a series of saccharides is shown in Fig. 4. Galactose and *N*-acetylgalactosamine did not block gp120 binding to clone 11-expressing COS cells. Mannan was the most potent inhibitor (IC_{50}, 6 μg/ml), followed by L-fucose (K_i, 6 mM); α-methyl D-mannoside (K_i, 15 mM), D-mannose (K_i, 23 mM); and *N*-acetylglucosamine (K_i, 70 mM). Human IgE, sialic acid, and mannose 6-phosphate had no effect on binding. As expected for a C-type lectin, the binding of gp120 to clone 11 required calcium and was blocked by EGTA (K_i, 0.3 mM). None of these sugars affected gp120 binding to CD4. Immune serum from an HIV-infected donor did block gp120/CD4 binding but not binding associated with the gp120-binding protein (data not shown).

Membrane Expression and Orientation of the gp120-Binding Protein. To provide additional evidence for the type II membrane orientation predicted by the cDNA sequence, the gp120-binding protein was expressed in HeLa cells by transfection with clone 11 cDNA ligated in the vector pcDNAI/Neo. After Geneticin selection, a high-binding population was enriched for by sterile sorting on a flow cytometer following staining with directly or indirectly fluorescein-conjugated gp120. Repeated sterile sorting after culture expansion resulted in a population of cells showing stable expression of the gp120-binding activity and with a growth phenotype indistinguishable from the parental, untransfected line. No evidence of cell aggregation was found, suggesting that the expressed lectin was not recognizing glycoproteins resident on the surface of adjacent HeLa cells. Following extended passage, the cells still bound high levels of vgp120 in a mannan-inhibitable manner (Fig. 5*A*).

Immunoprecipitation analyses revealed that the gp120-binding protein expressed on the transfected HeLa cell surface had a molecular mass of ≈46 kDa (data not shown), consistent with size predicted from the cDNA sequence. Flow cytometry studies using rabbit antiserum to the C-terminal peptide 564A confirmed the type II cell surface orientation of the gp120-binding protein on the transfected HeLa cells (Fig. 5*B*). No staining was seen with preimmune serum or with untransfected HeLa cells.

DISCUSSION

A placental library was chosen as the source of cDNA for screening by expression cloning because placental membranes, like neural tissue, bind gp120 in a CD4-independent manner.

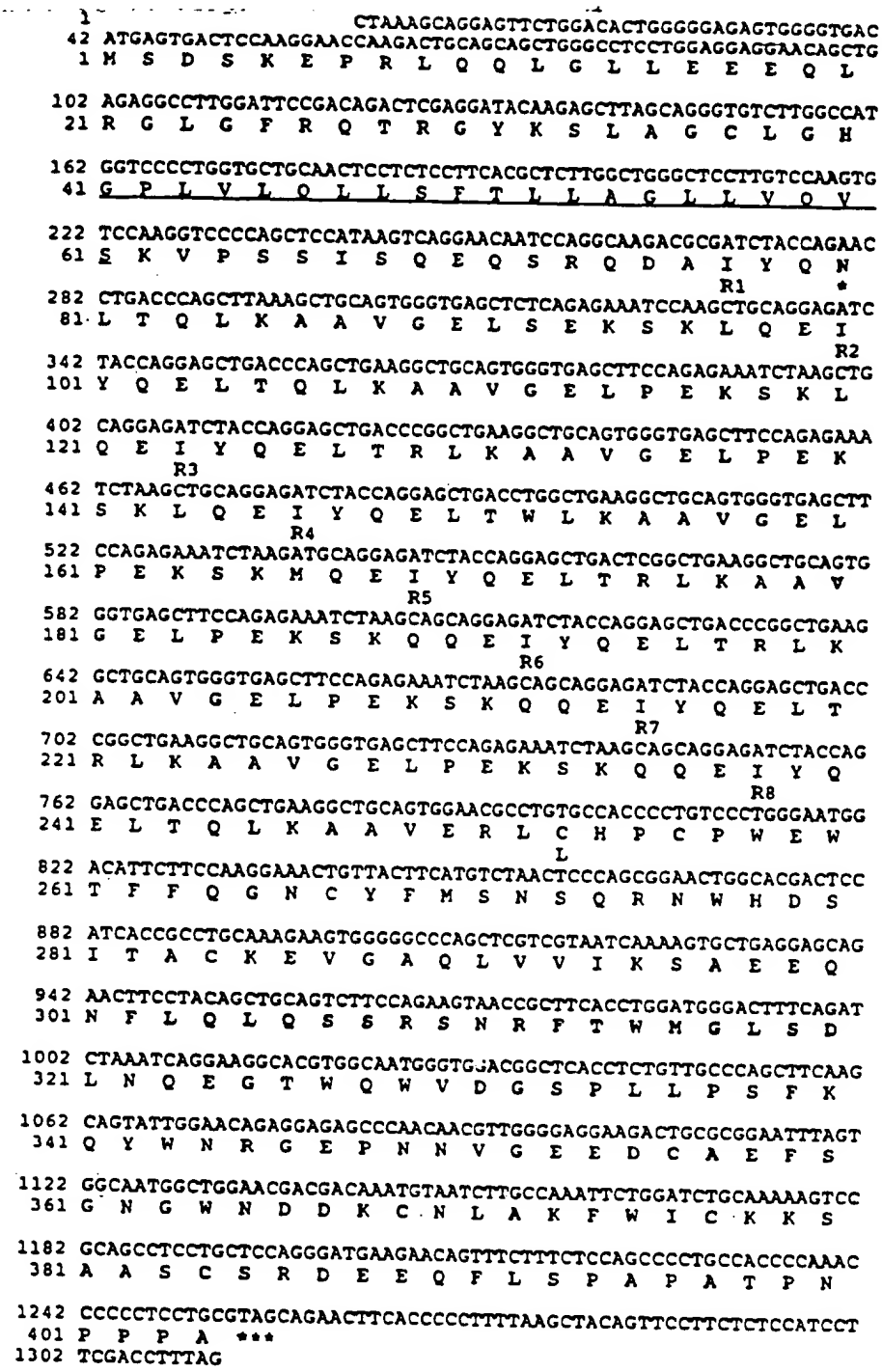


FIG. 2. Nucleotide and deduced amino acid sequence of clone 11 cDNA. The nucleotide sequence preceding the first ATG agrees with the Kozak consensus for translation initiation. The membrane-spanning sequence is underlined and the potential N-linked glycosylation site is marked by a star. The starts of the seven complete and eighth partial repeat (R1–R8) and the beginning of the lectin domain (L) are indicated.

The cDNA isolated encodes a 404-amino acid protein organized into three distinct domains. The sequence predicts a type II membrane orientation (N-terminal cytoplasmic) as suggested by the apparent absence of a signal sequence and the presence of a hydrophobic stop/transfer or anchor sequence (Gly^{41}–Ser^{61}) in the first domain. Application of the "positive-inside rule" (20) for the sequence within 15 residues of the hydrophobic region also predicts a cytoplasmic N terminus in agreement with the homology to membrane-associated C-type lectins with similar membrane orientation (27) (Fig. 3*B*). In addition, the reactivity with stably transfected cells of antiserum to the C-terminal peptide 564A supports this orientation (Fig. 5*B*).

The second domain consists of seven complete and one partial tandem repeat. Circular dichroism spectra in 40% trifluoroethanol of a consensus repeat peptide beginning with

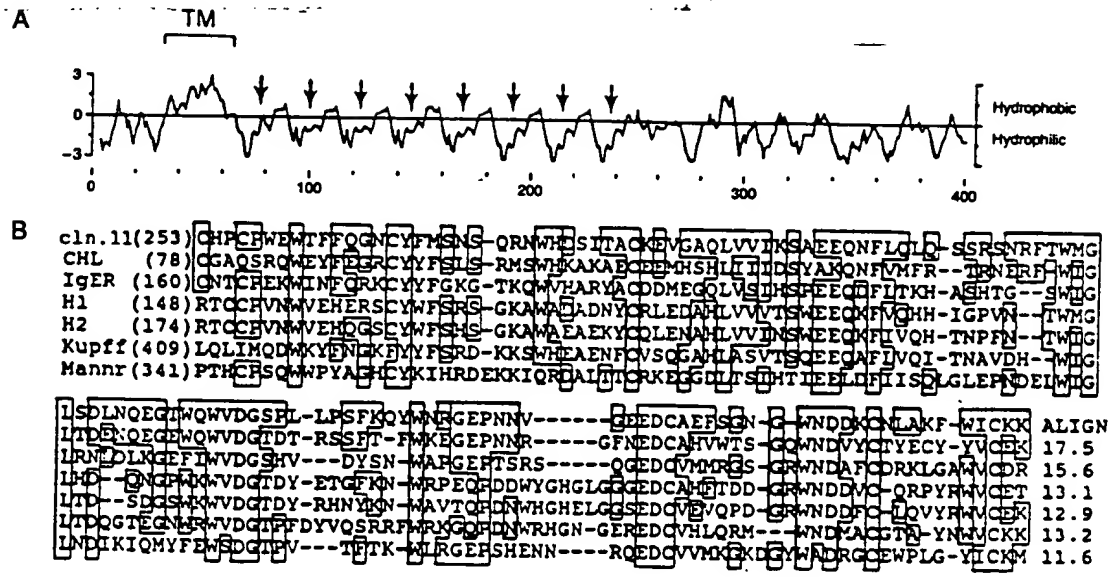


FIG. 3. (*A*) Hydropathicity plot. The predicted transmembrane segment (TM) is bracketed and the starts of the eight amphipathic repeats are indicated by arrows. (*B*) Amino acid alignment of the clone 11 C-type lectin domain. Residues identical to the gp120-binding protein (clone 11) are boxed. ALIGN scores > 3.0 indicate significant sequence similarity. CHL, chicken hepatic lectin; IgER, low-affinity IgE receptor; H1 and H2, human asialoglycoprotein receptors; Kupff, rat Kupffer cell receptor; Mannr, human macrophage mannose receptor.

the β-turn PEKSKLQEIYQELTQLKAAVGEL (single-letter amino acid code) demonstrated an all-α-helical structure (data not shown). Homology to other repeat domains suggested possible tertiary structures including antiparallel helix bundles or a multimeric parallel helix bundle, which would function as spacers to separate the lectin domain from the membrane.

The third domain shows homology to other C-type lectins and contains the conserved motif Trp-Asn-Asp, typical of this group (25). As shown in Fig. 3*B*, the most closely related sequences were the group of type II membrane protein C-type lectins: chick hepatic lectin (21), low-affinity IgE receptor (22), the asialoglycoprotein receptors (23), and the rat Kupffer cell receptor (24). The most similar mannose-binding lectin was one of the eight carbohydrate-recognition domains of the human macrophage mannose receptor (25, 26).

Despite the higher homology to lectins that bind terminal galactose and *N*-acetylglucosamine/galactosamine (27), inhibition studies using sugars and purified gp120 suggest that the terminal mannose residues of high-mannose chains are the primary determinants of binding. For these experiments three forms of gp120 were used: bgp120, which contains only high-mannose structures (28), and vgp120 and ngp120, which contain high-mannose and complex forms (29–31). All three forms have terminal mannose residues in common and all bound with similar affinity (Fig. 1*A*).

A number of studies have pointed to the importance of HIV envelope oligosaccharide side chains (32–35) and, specifically, mannose residues (36–38) in viral infectivity and syncytium formation. The high-affinity recognition of these residues by cell-associated mannose-binding lectins also predicts that such side chains may play a significant role in CD4-independent gp120 binding. Since the affinity of the mannose-binding protein for gp120 exceeds that of CD4, lectins of this type would be effective competitors for gp120 and viral binding on those cells that also express CD4.

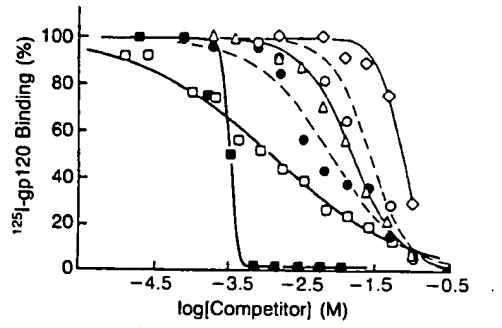


FIG. 4. Inhibition of gp120 binding to COS cells expressing the gp120-binding protein. Both ngp120 (open symbols) and vgp120 (filled symbols) were used and the relative values were the same with both forms of gp120. Mannan concentration is expressed as mg/ml. □, Mannan (IC_{50}, 6 μg/ml); ●, L-fucose (K_i, 6 mM); △, α-methyl D-mannoside (K_i, 15 mM); ○, D-mannose (K_i, 23 mM); ◇, *N*-acetylglucosamine (K_i, 70 mM); ■, EGTA (K_i, 0.3 mM).

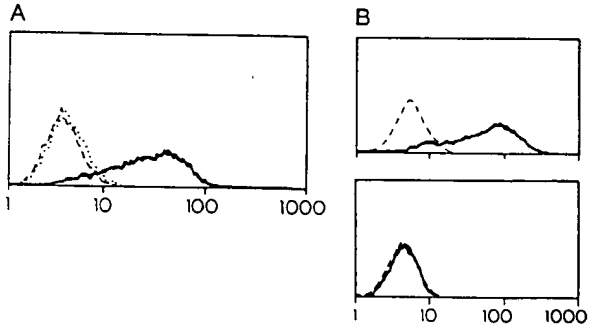


FIG. 5. Stable expression of gp120-binding protein on HeLa cells. HeLa cells were transfected with pcDNAI/Neo vector containing clone 11 cDNA, selected with Geneticin, and sterile-sorted by flow cytometry to enrich for high expression. (*A*) Flow cytometric analysis of transfected HeLa cells incubated with vgp120 (100 nM) (——), with buffer (- - -), or with mannan (4 mg/ml) followed by vgp120 (100 nM) (· · · ·). vgp120 binding was detected by antibody 110-4 followed by a fluorescein-labeled anti-mouse Ig reagent. (*B*) Reactivity of clone 11-transfected (*Upper*) or control (*Lower*) HeLa cells with rabbit preimmune serum (- - -) or antiserum to peptide 564A (——). Ordinate, cell number per channel; abscissa, log green-channel fluorescence.

Both the mannose-specific plant lectins (32, 34, 36, 38) and the human serum 32-kDa mannose-binding protein (39) can inhibit infection of T cells by HIV by a mechanism that does not appear to substantially disrupt gp120/CD4 interactions (38, 40). The consequences of virus binding to a membrane-associated mannose-binding protein are not known, however, and could include CD4-independent infection, as has been suggested in macrophages (40), or entry of the virus into an endosomal pathway and inactivation in the lysosomal compartment, or as seen in epithelial cells, transcytosis (14, 41). Preliminary HIV infection studies on clone 11 transfected HeLa cells are consistent with a role of this lectin in virus binding and internalization, but not infection of these cells (data not shown).

Mannose-binding proteins appear to be able to discriminate between the carbohydrate structures present on gp120 and those present on the surface of normal cells. Since glycosylation of gp120 is directed by host cellular enzymes, this suggests that control of normal cellular glycosylation mechanisms is disrupted by HIV infection. The ability to differentiate viral from host cell oligosaccharides raises the possibility of a therapeutic role for mannose-binding proteins in HIV infection.

We thank Brian Seed for the gift of the pCDM8 placental library and πH3MCD4, Mike Neubauer for DNA sequencing, Jan McClure for ngp120, Jim Blake for peptide synthesis, Greg Plowman for human mRNA, Jeff Ledbetter for G17-2 antibody, Bruce Travis for constructing the v11K120 recombinant vaccinia virus, Sridhar Pennathur for gp120 production, and Derek Hewgill for flow cytometry assistance. We thank Sandro Aruffo for critical review of the manuscript, and Shiu-Lok Hu and Mohammed Shoyab for continued support.

1. Sattentau, Q. J. & Weiss, R. A. (1988) *Cell* **52**, 631–633.
2. Clapham, P. R., Weber, J. N., Whitby, D., McIntosh, K., Dalgleish, A. G., Maddon, P. J., Deen, K. C., Sweet, R. W. & Weiss, R. A. (1989) *Nature (London)* **337**, 368–370.
3. Harouse, J. M., Kunsch, C., Hartle, H. T., Laughlin, M. A., Hoxie, J. A., Wigdahl, B. & Gonzales-Scarano, F. (1989) *J. Virol.* **63**, 2527–2533.
4. Weber, J., Clapham, P., McKeating, J., Stratton, M., Robey, E. & Weiss, R. (1989) *J. Gen. Virol.* **70**, 2653–2660.
5. Kunsch, C., Hartle, H. T. & Wigdahl, B. (1989) *J. Virol.* **63**, 5054–5061.
6. Cao, Y., Friedman-Kien, A. E., Huang, Y., Li, X. I., Mirabile, M., Moudgil, T., Zucker-Franklin, D. & Ho, D. D. (1990) *J. Virol.* **64**, 2553–2559.
7. Omary, M. B., Brenner, D. A., de Grandpre, L. Y., Roebuck, K. A., Richman, D. D. & Kagnoff, M. F. (1991) *AIDS* **5**, 275–281.
8. Kozlowski, M. R., Sandler, P., Lin, P.-F. & Watson, A. (1991) *Brain Res.* **553**, 300–304.
9. Dreyer, E. B., Kaiser, P. K., Offermann, J. T. & Lipton, S. A. (1990) *Science* **248**, 364–367.
10. Brenneman, D. E., Westbrook, G. L., Fitzgerald, S. P., Ennist, D. L., Elkins, K. L., Ruff, M. R. & Pert, C. B. (1988) *Nature (London)* **335**, 639–642.
11. Gearing, D. P., King, J. A., Gough, N. M. & Nicola, N. A. (1989) *EMBO J.* **8**, 3667–3676.
12. Larsen, A., Davis, T., Curtis, B. M., Gimpel, S., Sims, J. E., Cosman, D., Park, L., Sorenson, E., March, C. J. & Smith, C. A. (1990) *J. Exp. Med.* **172**, 1559–1570.
13. Hansen, J.-E. S., Clausen, H., Hu, S.-L., Nielsen, J. O. & Olofsson, S. *Arch. Virol.*, in press.
14. Bourinbaiar, A. S. & Phillips, D. M. (1991) *J. Acquired Immune Defic. Syndr.* **4**, 56–63.
15. Linsley, P. S., Ledbetter, J. A., Kinney-Thomas, E. & Hu, S.-L. (1988) *J. Virol.* **62**, 3695–3702.
16. Peterson, A. & Seed, B. (1988) *Cell* **54**, 65–72.
17. Kozlowski, M. R. & Watson, A. (1990) *Antiviral. Chem. Chemother.* **1**, 175–182.
18. Curtis, B. M., Widmer, M. B., deRoos, P. & Qwarnstrom, E. E. (1990) *J. Immunol.* **44**, 1295–1303.
19. Schnittman, S. M., Lane, H. C., Roth, J., Burrows, A., Folks, T. M., Kehrl, J. H., Koenig, S., Berman, P. & Fauci, A. S. (1988) *J. Immunol.* **141**, 4181–4186.
20. von Heijne, G. & Gavel, Y. (1988) *Eur. J. Biochem.* **174**, 671–678.
21. Drickamer, K. (1981) *J. Biol. Chem.* **256**, 5827–5839.
22. Kikutani, H., Inui, S., Sato, R., Barsumian, E. L., Owaki, H., Yamasaki, K., Kaisho, T., Uchibayashi, N., Hardy, R. R., Hirano, T., Tsunasawa, S., Sakiyama, F., Suemura, M. & Kishimoto, T. (1986) *Cell* **47**, 657–665.
23. Spiess, M. & Lodish, H. F. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6465–6469.
24. Hoyle, G. W. & Hill, R. L. (1988) *J. Biol. Chem.* **263**, 7487–7492.
25. Ezekowitz, R. A. B., Sastry, K., Bailly, P. & Warner, A. (1990) *J. Exp. Med.* **172**, 1785–1794.
26. Taylor, M. E., Conary, J. T., Lennartz, M. R., Stahl, P. D. & Drickamer, K. (1990) *J. Biol. Chem.* **265**, 12156–12162.
27. Drickamer, K. (1989) *Ciba Found. Symp.* **145**, 45–61.
28. Hsieh, P. & Robbins, P. W. (1984) *J. Biol. Chem.* **259**, 2375–2382.
29. Mizuochi, T., Spellman, M. W., Larkin, M., Solomon, J., Basa, L. J. & Feizi, T. (1988) *Biochem. J.* **254**, 599–603.
30. Mizuochi, T., Matthews, T. J., Kato, M., Hamako, J., Titani, K., Solomon, J. & Feizi, T. (1990) *J. Biol. Chem.* **265**, 8519–8524.
31. Geyer, H., Holschbach, C., Hunsmann, G. & Schneider, J. (1988) *J. Biol. Chem.* **263**, 11760–11767.
32. Lifson, J., Coutre, S., Huang, E. & Engleman, E. (1986) *J. Exp. Med.* **164**, 2101–2106.
33. Matthews, T. J., Weinhold, K. J., Lyerly, H. K., Langlois, A. J., Wigzell, H. & Bolognesi, D. P. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5424–5428.
34. Hansen, J.-E. S., Nielsen, C. M., Nielsen, C., Heegaard, P., Mathiesen, L. R. & Nielsen, J. O. (1989) *AIDS* **3**, 635–641.
35. Fenouillet, E., Gluckman, J. C. & Bahraoui, E. (1990) *J. Virol.* **64**, 2841–2848.
36. Robinson, W. E., Jr., Montefiori, D. C. & Mitchell, W. M. (1987) *AIDS Res. Hum. Retroviruses* **3**, 265–282.
37. Muller, W. E. G., Schroder, H. C., Reuter, P., Maidhof, A., Uhlenbruck, G. & Winkler, I. (1990) *AIDS* **4**, 159–162.
38. Gattegno, L., Ramdani, A., Jouault, T., Saffar, L. & Gluckman, J.-C. (1992) *AIDS Res. Hum. Retroviruses* **8**, 27–37.
39. Ezekowitz, R. A. B., Kuhlman, M., Groopman, J. E. & Byrn, R. A. (1989) *J. Exp. Med.* **169**, 185–196.
40. Larkin, M., Childs, R. A., Matthews, T. J., Thiel, S., Mizuochi, T., Lawson, A. M., Savill, J. S., Haslett, C., Diaz, R. & Feizi, T. (1989) *AIDS* **3**, 793–798.
41. Amerongen, H. M., Weltzin, R., Farnet, C. M., Michetti, P., Haseltine, W. A. & Neutra, M. R. (1991) *J. Acquired Immune Defic. Syndr.* **4**, 760–765.

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 8212–8216, July 1999
Neurobiology

Chemokines and activated macrophages in HIV gp120-induced neuronal apoptosis

MARCUS KAUL AND STUART A. LIPTON*

CNS Research Institute, Brigham and Women's Hospital, and Program in Neuroscience, Harvard Medical School, Boston, MA 02115

Edited by Solomon H. Snyder, Johns Hopkins University School of Medicine, Baltimore, MD, and approved May 3, 1999 (received for review December 18, 1998)

ABSTRACT **HIV-1 glycoprotein gp120 induces injury and apoptosis in rodent and human neurons *in vitro* and *in vivo* and is therefore thought to contribute to HIV-associated dementia. In addition to CD4, different gp120 isolates bind to the α- or β-chemokine receptors CXCR4 and CCR5, respectively. These and other chemokine receptors are on brain macrophages/microglia, astrocytes, and neurons. Thus, apoptosis could occur via direct interaction of gp120 with neurons, indirectly via stimulation of glia to release neurotoxic factors, or via both pathways. Here we show in rat cerebrocortical cultures that recapitulate the type and proportion of cells normally found in brain, i.e., neurons, astrocytes, and macrophages/microglia, that the β-chemokines RANTES (regulated on activation, normal T cell expressed and secreted) and macrophage inflammatory protein (MIP-1β) protect neurons from $gp120_{SF2}$-induced apoptosis. The $gp120_{SF2}$ isolate prefers binding to CXCR4 receptors, similar to the physiological α-chemokine ligands, stromal cell-derived factor (SDF)-1α/β. SDF-1α/β failed to prevent $gp120_{SF2}$ neurotoxicity, and in fact also induced neuronal apoptosis. We could completely abrogate $gp120_{SF2}$-induced neuronal apoptosis with the tripeptide TKP, which inhibits activation of macrophages/microglia. In contrast, TKP or depletion of macrophages/microglia did not prevent SDF-1 neurotoxicity. Inhibition of p38 mitogen-activated protein kinase ameliorated both $gp120_{SF2}$- and SDF-1-induced neuronal apoptosis. Taken together, these results suggest that $gp120_{SF2}$ and SDF-1 differ in the cell type on which they stimulate CXCR4 to induce neuronal apoptosis, but both ligands use the p38 mitogen-activated protein kinase pathway for death signaling. Moreover, $gp120_{SF2}$-induced neuronal apoptosis depends predominantly on an indirect pathway via activation of chemokine receptors on macrophages/microglia, whereas SDF-1 may act directly on neurons or astrocytes.**

About half of children and a quarter of adults infected with HIV-1 eventually develop dementia (1). Transgenic mice expressing the HIV-1 envelope glycoprotein gp120 manifest neuropathological features that resemble in many ways the findings in brains of AIDS patients (2). *In vitro* and *in vivo*, gp120 produces injury and apoptosis in both primary rodent and human neurons (3–9). Recent evidence has shown that gp120 binds, respectively, to macrophages and T cells via the chemokine receptors CCR5 and CXCR4, which, in addition to CD4, function as coreceptors for HIV-1 (10–13). Nonetheless, CCR5 and CXCR4, as well as other chemokine receptors, are also present on neurons and astrocytes (12, 14–16). Thus, a major question addressed in the present study is whether gp120-induced neuronal injury occurs as a consequence of direct interaction with neurons via chemokine receptors and their cognate G protein-signaling systems (13, 17) or indirectly via the release of macrophage toxic factors, as previously suggested from *in vitro* experiments with gp120-conditioned medium after macrophage depletion (18–22). Finally, both direct and indirect pathways in conjunction could contribute to neuronal death, in a manner similar to that recently shown for the CXCR4-mediated killing of $CD8^+$ T cells (23). Although some gp120 variants can signal via chemokine receptors on neuronal cell lines and on isolated rodent neurons (13, 17), the importance of cell–cell interactions in the brain mandates that disease pathogenesis *in vitro* be approached in a culture system that recapitulates the type and proportion of cells normally found in brain, i.e., neurons, astrocytes, and macrophages/microglia.

Here we show in such a "mixed" culture system (24) that the β-chemokines RANTES (regulated on activation, normal T cell expressed and secreted) or MIP-1β can protect rat cerebrocortical neurons from gp120-induced apoptosis, whereas the α-chemokines SDF-1α and β not only fail to prevent gp120 neurotoxicity but induce neuronal apoptosis themselves. The tuftsin-derived tripeptide TKP (Thr-Lys-Pro), which inhibits macrophage/microglial activation (25–28), completely abrogates gp120-induced neuronal apoptosis. In contrast, TKP or depletion of monocytoid cells from the culture does not prevent the neurotoxicity of SDF-1, indicating that it is independent of macrophages/microglia. However, inhibition of the p38 mitogen-activated protein kinase (MAPK) signaling pathway ameliorates both gp120- and SDF-1-induced neuronal damage. Thus, $gp120_{SF2}$ and SDF-1 stimulate CXCR4 receptors on different cell types; yet in both cases, p38 MAPK is in the signaling pathway to neuronal apoptosis. Additionally, our results suggest that $gp120_{SF2}$-induced neuronal apoptosis is mediated indirectly via chemokine receptors on macrophages/microglia, whereas the α-chemokines SDF-1α and β appear to exert their action directly on neurons or astrocytes.

MATERIALS AND METHODS

Peptides and Recombinant Proteins. The tripeptide TKP (Thr-Lys-Pro; tuftsin fragment 1–3) was obtained from Sigma. Recombinant human MIP-1β, SDF-1α, SDF-1β, and recombinant rat RANTES were purchased from R&D Systems and Endogen (Cambridge, MA), respectively. HIV-1 envelope glycoprotein gp120 from the strain SF2 was obtained from the National Institutes of Health AIDS Research and Reference Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health (29). Additional gp120s from HIV-1 strains IIIB and RF2 were obtained from Genentech and the National Cancer Institute,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: RANTES, regulated on activation, normal T cell expressed and secreted; MAPK, mitogen-activated protein kinase; MAP-2, microtubule-associated protein-2; NMDA, *N*-methyl-D-aspartate; TKP, Thr-Lys-Pro; MIP, macrophage inflammatory protein; SDF, stromal cell-derived factor.
*To whom reprint requests should be addressed. e-mail: slipton@rics.bwh.harvard.edu.

respectively, and in previous experiments were found to produce neurotoxicity similar to $gp120_{SF2}$ (4, 18, 21, 30–33). Tumor necrosis factor α, IFN-γ, and IL-1β were from Genzyme, GIBCO/BRL, and Endogen (Cambridge, MA), respectively.

Preparation and Treatment of Rat Cerebrocortical Cultures. Cerebrocortical cultures were prepared from embryos of Sprague–Dawley rats at day 15–17 of gestation, as described (34, 35). Cultures were used for experiments after 17–24 days in culture. These cultures contain neurons, astrocytes, and macrophages/microglia, as determined with specific immunolabeling. Before some experiments, macrophages/microglia or neurons were depleted from the cultures by exposure to 7.5 mM L-leucine methyl ester or 2 mM *N*-methyl-D-aspartate (NMDA), respectively (18, 34). Absence of macrophages/microglia or neurons in these cultures was confirmed immunocytochemically, by using antibodies to ED-1 and microtubule-associated protein-2 (MAP-2), respectively. In some experiments, the Griess reaction was used to measure nitrite levels in the culture medium as an index of NO release (36).

Incubation of Cells with TKP, Chemokines, and p38 MAPK Inhibitor. Cultures were transferred into Earle's balanced salt solution and incubated for 24 hr with gp120, chemokines, TKP, p38 MAPK inhibitor SB203580 (Calbiochem), or combinations thereof. Chemokines or TKP were applied for 5 min and SB203580 for 15 min before gp120 exposure.

Assessment of Neuronal Apoptosis. Apoptosis in these cultures was assessed by using multiple methods with concordant results as detailed (24). We routinely used a combination of staining of permeabilized cells with propidium iodide to determine apoptotic morphology and a neuron-specific antibody to identify cell type. In brief, cells were fixed for 5 min with ice-cold acetone at −20°C and, after three washes in PBS, for 4 min with 2% (wt/vol) paraformaldehyde solution in PBS at room temperature. Acetone-paraformaldehyde-fixed cells were permeabilized by using 0.2% Tween 20/PBS, and nonspecific binding sites were blocked by incubation for 1 hr with a 10% solution of heat-inactivated goat serum in 0.2% Tween 20/PBS. To specifically stain neurons, cells were then incubated for 4 hr at room temperature or overnight at 4°C with 1:500 dilutions of anti-MAP-2 (Sigma) or anti-NeuN mAb (Chemicon). Their respective nonspecific isotype antibodies served as controls. After three washes, the cells were incubated in a secondary polyclonal antibody conjugated either to FITC or to horseradish peroxidase. In the case of horseradish peroxidase-coupled polyclonal antibody, diaminobenzidine served as the color substrate developed by incubation in a mixture of 1 mg/ml diaminobenzidine and 0.8% H_2O_2 at a ratio of 3:1. Cellular nuclei were subsequently stained with 20 μg/ml propidium iodide for 5 min in the dark, and then coverslips were mounted on glass slides. Experiments were replicated at least three times, with triplicate values in each experiment. Statistical significance was determined by using ANOVA followed by a Scheffé or Bonferroni/Dunn post hoc test.

RESULTS AND DISCUSSION

We scored the number of apoptotic cerebrocortical neurons in culture exposed to $gp120_{SF2}$ by using a combination of propidium iodide staining of permeabilized cells to identify apoptotic nuclei and MAP-2 or NeuN immunostaining to specifically identify neurons (Fig. 1). Additional experiments, by using glial fibrillary acidic protein antibody to identify astrocytes in mixed neuronal/glial cultures or cultures depleted of neurons by prior exposure to NMDA, revealed no significant apoptosis in glial cells after $gp120_{SF2}$ exposure under our culture conditions (data not shown). The β-chemokines, RANTES and MIP-1β (each at 20 nM), abrogated neuronal apoptosis induced by 200 pM recombinant $gp120_{SF2}$ (Fig. 2), whereas BSA (0.001% = 144 nM) and the α-chemokines SDF-1α or SDF-1β (20–50 nM) did not protect. In fact, these α-chemokines produced neurotoxicity on their own (≈2-fold increase in neuronal apoptosis compared with control, Fig. 3). MIP-1β and RANTES presumably inhibit the neurotoxic effect of $gp120_{SF2}$ in an indirect manner, because RANTES binds to the β-chemokine receptors CCR1, CCR3, and CCR5, and MIP-1β binds CCR5 (or a functional rat homologue) (37–39), whereas $gp120_{SF2}$ (and SDF-1α/β) interact with the α-chemokine receptor CXCR4 (40). Note that although $gp120_{SF2}$ may also interact to a lesser degree with the β-chemokine receptor CCR5 on some transfected cell lines (41), this has not been shown to occur on primary cells, as used here. In line with these results with $gp120_{SF2}$, X4 (CXCR4-preferring) virus or dual tropic (X4/R5) virus was recently shown to cause neuronal apoptosis in human cerebrocortical cell cultures (42). Importantly, rodent cerebrocortical cultures are a suitable model system to study these actions of gp120 because these species express CXCR4 homologues that, like the human CXCR4, are capable of mediating HIV-1 infection via gp120 binding (43, 44). Previously, we found in our rodent cultures that gp120-induced neuronal damage was prevented by anti-gp120 antibodies but not by anti-CD4 antibodies, proving the specificity of the effect of gp120 but also implying that CD4 was not necessary for neurotoxicity (4, 30).

However, these results with RANTES and MIP-1β do not tell us whether the neuroprotective effect of these β-chemokines and, for that matter, the neurotoxic effect of gp120 is mediated by macrophages, astrocytes, neurons, or by simultaneous action on two or all three cell types. To address this query, we used the macrophage inhibitory tripeptide Thr-Lys-Pro (TKP), which has been shown to specifically prevent activation of macrophages/microglia and subsequent release of their toxic factors both *in vitro* and *in vivo*, whereas control peptides have no effect (25–28). TKP is comprised of three of the four amino acid residues of tuftsin, a well characterized peptide known to display the opposite effect, i.e., activation of macrophages (45). In our experiments, TKP (50 μM) protected neurons from gp120-induced apoptosis (Fig. 4*A*), similar to our previous experiments with macrophages depleted from the cultures (18). In contrast, SDF-1β-induced neuronal

FIG. 1. Immunofluorescence images of neurons labeled with MAP-2 (*Left*) and propidium iodide (*Center*) to show apoptosis of neurons (*Right*, superimposed images) in mixed neuronal/glial cultures after overnight exposure to 200 pM gp120. After propidium iodide labeling, apoptotic neurons appear small, round, and intensely fluorescent (indicated with white arrows in the *Center* image).

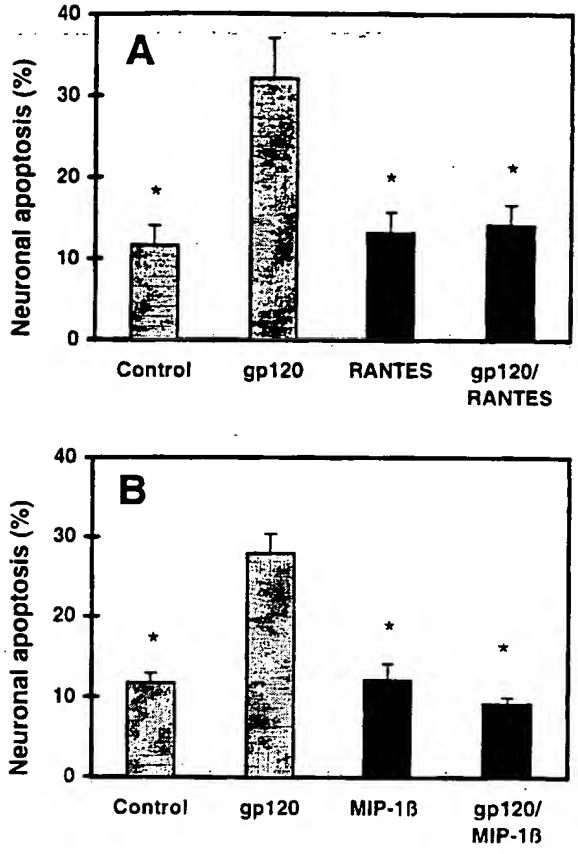


FIG. 2. Protection of rat cerebrocortical neurons from gp120-induced apoptosis by the β-chemokines RANTES and MIP-1β. *(A)* Neuroprotection by recombinant rat RANTES (20 nM). *(B)* Neuroprotection by recombinant human MIP-1β (20 nM). Rat cerebrocortical cultures were incubated for 24 hr with or without 200 pM recombinant gp120 and in the presence or absence of each chemokine. After fixation and permeabilization, neurons were identified by immunostaining for MAP-2 or NeuN, and apoptotic cells were assessed by propidium iodide staining. *, $P < 0.01$ compared with value for gp120.

apoptosis was not abrogated by TKP (Fig. 4*B*) and also occurred in cultures depleted of macrophages/microglia (data not shown).

Several lines of evidence confirmed prior reports that TKP exerted its effect specifically on macrophages/microglia and not on astrocytes or neurons (25–28). For example, in the absence of macrophages/microglia, TKP did not inhibit NO release by cytokine-activated astrocytes (Fig. 4*C*), and TKP did not interfere with NMDA-induced neuronal apoptosis (ref. 24; data not shown). These findings indicate that activated macrophages are necessary for $gp120_{SF2}$- but not SDF-1-induced neuronal apoptosis if the various types and proportion of cells present in the brain are also present in the culture system. This fact does not exclude a direct interaction of gp120 with neuronal or astroglial CXCR4 or other chemokine receptors. But if this interaction occurs, in contrast to the effect of SDF-1, it is apparently not sufficient to trigger neuronal apoptosis in these mixed neuronal/glial cultures.

The number of HIV-1-infected cells in the brain is relatively small, and productively infected cells are exclusively of monocytoid lineage (reviewed in ref. 1). This observation suggests that HIV-1 initiates a neurodegenerative process that entails amplification to produce pronounced central nervous system injury (1). Indeed, in culture systems of both rodent and human brain, HIV-1-infected or gp120-stimulated macrophages and microglia have been found to release neurotoxins that contribute to the neurodegenerative process, at least in part, by excessive stimulation of the NMDA subtype of glutamate receptor (1). The fact that gp120-transgenic mice manifest neuronal damage resembling that found both in rodent cultures and in human brain with HIV-associated dementia indicates that, even in the absence of intact HIV-1, a fragment of the virus is sufficient to trigger important aspects of this amplification cascade in the neurodegenerative process in our *in vitro* system, which therefore has relevance to *in vivo* pathogenicity.

In another series of experiments, we tested a variety of inhibitors of intracellular signaling cascades for their ability to prevent neuronal apoptosis associated with gp120. These inhibitors included PD98059 (2 μM) to inhibit extracellular regulated kinase MAPK, pyrrolidine dithiocarbamate (PDTC, 5 μM) to inhibit NF-κB, and SB203580 (10 μM) to specifically inhibit p38 MAPK (46). Of these, only SB203580 substantially attenuated $gp120_{SF2}$-induced neuronal apoptosis (Fig. 5*A*), implicating the p38 MAPK pathway in gp120-activated death signaling. Inhibition of p38 MAPK also ameliorated SDF-1 neurotoxicity (Fig. 5*B*). This finding indicates that the neurotoxic processes initiated by $gp120_{SF2}$ and SDF-1 use the common MAPK signaling pathway involving p38. Because SDF-1-induced neurotoxicity occurs in the virtual absence of macrophages/microglia, p38 MAPK must be activated as a stress response in neurons or astrocytes. In fact, from previous

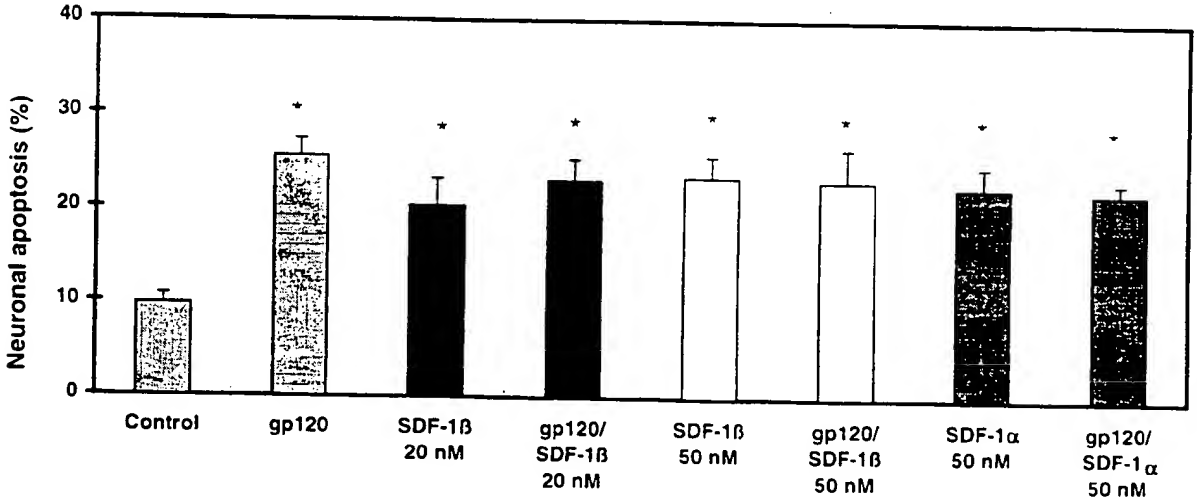


FIG. 3. Neurotoxic effect of SDF-1α and β and lack of protection from gp120-induced apoptosis. Treatment, identification, and analysis of cells as in the legend to Fig. 2. *, $P < 0.01$ compared with value for control but not significantly different from each other.

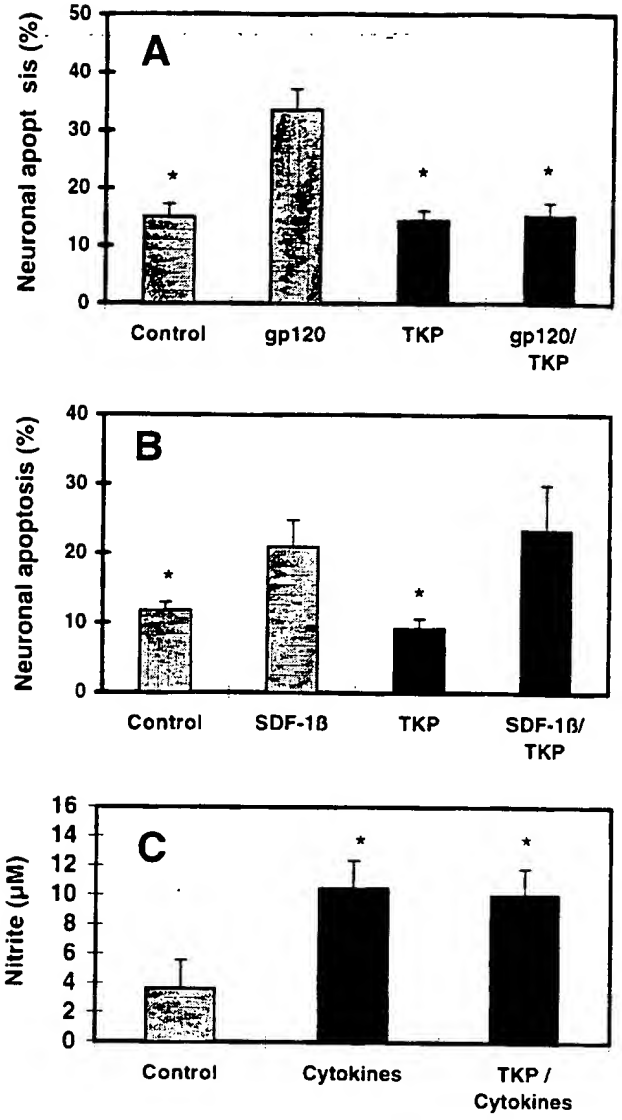


FIG. 4. Effects of macrophage-inhibitory peptide TKP on gp120- and SDF-1-induced neuronal apoptosis and on astrocyte activation of immunologic NO synthase. *(A)* TKP protected neurons from $gp120_{SF2}$ toxicity. *(B)* TKP did not protect neurons from SDF-1β toxicity. Experimental conditions and analysis of apoptosis as in the legend to Fig. 2, except TKP (50 μM) was used instead of β-chemokines. *, $P < 0.01$ compared with value for gp120 or SDF-1β. *(C)* As a control to show that TKP did not prevent astrocyte activation, TKP did not inhibit release of NO from cytokine-stimulated astrocytes in cultures depleted of macrophages/microglia (see *Materials and Methods*). Astrocytic iNOS was induced by treatment with the cytokines tumor necrosis factor α (200 units/ml), IFN-γ (200 units/ml), and IL-1β (1 ng/ml) for 24 hr in the presence or absence of TKP. "Control" indicates samples without TKP and cytokines. Nitrite levels were monitored in the culture medium as an index of NO release by astrocytes. *, $P < 0.01$ compared with value for control but not significantly different from each other.

work (47, 48), we know that excitotoxic (NMDA) receptor-mediated apoptosis in neurons is mediated, at least in part, by a p38 pathway, and we also know that gp120-induced neuronal damage is prevented by NMDA antagonists (31). Hence, a neuronal p38 pathway perforce must come into play in gp120-induced neurotoxicity. However, we cannot exclude the possibility that gp120 and SDF-1 also activate p38 in macrophages/microglia. In fact, this is likely to occur because activation of p38 MAPK has been reported in activated macrophages/microglia (46). Additionally, immunocytochemical experiments in our culture system have revealed activated (diphosporylated) p38 in both neurons and macrophages (data not shown).

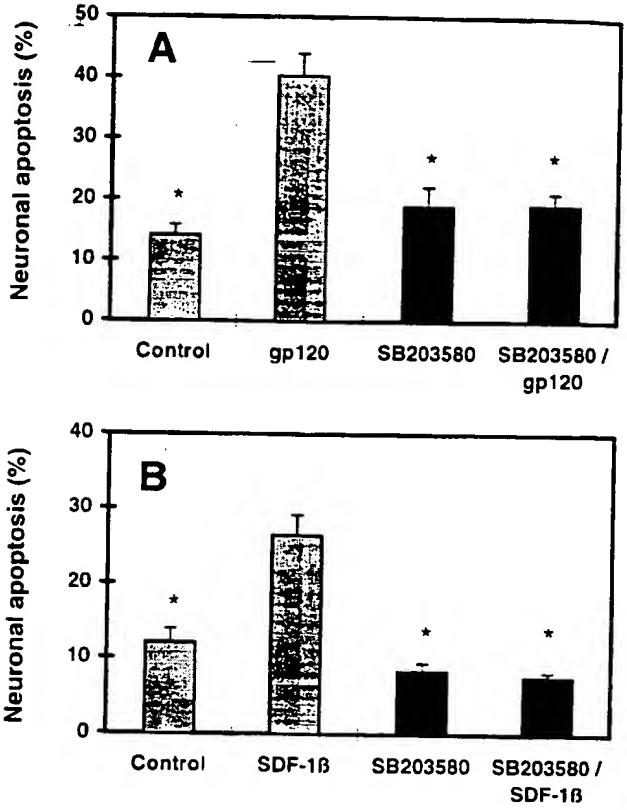


FIG. 5. Inhibition of p38 MAPK reduces gp120- and SDF-1-induced neuronal apoptosis. In the presence or absence of the p38 MAPK inhibitor SB203580 (10 μM), cerebrocortical cultures were incubated for 24 hr with or without 200 pM recombinant $gp120_{SF2}$ (*A*) or 20 nM SDF-1β (*B*). Treatment, identification, and analysis of neurons as in the legend to Fig. 2. *, $P < 0.01$ compared with value for gp120 or SDF-1β.

Taken together, the simplest explanation of our findings with RANTES, MIP-1β, and TKP is that gp120 neurotoxicity depends predominantly on activation of chemokine receptors on macrophages and microglia rather than solely on neurons or astrocytes. In contrast, SDF-1-induced neuronal apoptosis does not require the activation or presence of macrophages/microglia, and therefore the pathophysiologically relevant stimulus for neuronal cell death from this α-chemokine may be transmitted via astrocytes or directly on neurons. Moreover, the fact that the neurotoxic effect of a T cell tropic strain of gp120 ($gp120_{SF2}$), which has been shown to signal via the α-chemokine receptor CXCR4 (40), can be offset by β-chemokines binding solely to CCR5 (MIP-1β) may indicate that there is a novel pattern of cross-talk between the signaling pathways of various G protein-coupled chemokine receptors. Additionally, although $gp120_{SF2}$ and SDF-1 differ in the cell type on which they stimulate CXCR4 to induce neuronal apoptosis, both ligands use the p38 MAPK pathway for death signaling. Such signaling cascades may offer new therapeutic targets for interrupting the indirect macrophage pathway to gp120-induced neuronal apoptosis as well as the nonmacrophage-mediated pathway to α-chemokine (SDF-1)-induced neuronal apoptosis.

This study was supported by a fellowship award from the Deutsche Forschungsgemeinschaft to M.K. and National Institutes of Health Grants R01 EY09024 and P01 HD29587 to S.A.L.

1. Lipton, S. A. & Gendelman, H. E. (1995) *N. Engl. J. Med.* **332,** 934–940.
2. Toggas, S. M., Masliah, E., Rockenstein, E. M., Rall, G. F., Abraham, C. R. & Mucke, L. (1994) *Nature (London)* **367,** 188–193.
3. Brenneman, D. E., Westbrook, G. L., Fitzgerald, S. P., Ennist, D. L., Elkins, K. L., Ruff, M. & Pert, C. B. (1988) *Nature (London)* **335,** 639–642.
4. Dreyer, E. B., Kaiser, P. K., Offermann, J. T. & Lipton, S. A. (1990) *Science* **248,** 364–367.
5. Müller, W. E. G., Schröder, H. C., Ushijima, H., Dapper, J. & Bormann, J. (1992) *Eur. J. Pharmacol.* **226,** 209–214.
6. Lannuzel, A., Lledo, P.-M., Lamghitnia, H. O., Vincent, J.-D. & Tardieu, M. (1995) *Eur. J. Neurosci.* **7,** 2285–2293.
7. Meucci, O. & Miller, R. J. (1996) *J. Neurosci.* **16,** 4080–4088.
8. Bagetta, G., Corasaniti, M. T., Malorni, W., Rainaldi, G., Berliocchi, L., Finazzi-Agrò, A. & Nisticò, G. (1996) *NeuroReport* **7,** 1722–1724.
9. Lannuzel, A., Barnier, J. V., Hery, C., Van Tan, H., Guibert, B., Gray, F., Vincent, J. D. & Tardieu, M. (1997) *Ann. Neurol.* **42,** 847–856.
10. Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., Gerard, C. & Sodroski, J. (1996) *Nature (London)* **384,** 179–183.
11. Clapham, P. R. (1997) *Trends Cell. Biol.* **7,** 264–268.
12. Hesselgesser, J., Halks-Miller, M., DelVecchio, V., Peiper, S. C., Hoxie, J., Kolson, D. L. & Horuk, R. (1997) *Curr. Biol.* **7,** 112–121.
13. Hesselgesser, J., Taub, D., Baskar, P., Greenberg, M., Hoxie, J., Kolson, D. L. & Horuk, R. (1998) *Curr. Biol.* **8,** 595–598.
14. Rottman, J. B., Ganley, K. P., Williams, K., Wu, L., Mackay, C. R. & Ringler, D. J. (1997) *Am. J. Pathol.* **151,** 1341–1350.
15. Lavi, E., Strizki, J. M., Ulrich, A. M., Zhang, W., Fu, L., Wang, Q., O'Connor, M., Hoxie, J. A. & González-Scarano, F. (1997) *Am. J. Pathol.* **151,** 1035–1042.
16. Heesen, M., Berman, M. A., Charest, A., Housman, D., Gerard, C. & Dorf, M. E. (1998) *Immunogenetics* **47,** 364–370.
17. Meucci, O., Fatatis, A., Simen, A. A., Bushnell, T. J., Gray, P. W. & Miller, R. J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14500–14505.
18. Lipton, S. A. (1992) *NeuroReport* **3,** 913–915.
19. Giulian, D., Wendt, E., Vaca, K. & Noonan, C. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 2769–2773.
20. Diop, A. G., Lesort, M., Esclaire, F., Sindou, P., Couratier, P. & Hugon, J. (1994) *Neurosci. Lett.* **165,** 187–190.
21. Lipton, S. A. (1994) *Neurosci. Res. Commun.* **15,** 31–37.
22. Flavin, M. P., Coughline, K. & Ho, L. T. (1997) *Neuroscience* **80,** 437–448.
23. Herbein, G., Mahlknecht, U., Batliwalla, F., Gregersen, P., Pappas, T., Butler, J., O'Brien, W. A., Verdin, E. (1998) *Nature (London)* **395,** 189–194.
24. Bonfoco, E., Krainc, D., Ankarcrona, M., Nicotera, P. & Lipton, S. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7162–7166.
25. Auriault, C., Joseph, M., Tartar, A. & Capron, A. (1983) *FEBS Lett.* **153,** 11–15.
26. Auriault, C., Joseph, M., Tartar, A., Bout, D., Tonnel, A. B. & Capron, A. (1985) *Int. J. Immunopharmacol.* **7,** 73–79.
27. Thanos, S., Mey, J. & Wild, M. (1993) *J. Neurosci.* **13,** 455–466.
28. Rogove, A. D. & Tsirka, S. E. (1998) *Curr. Biol.* **8,** 19–25.
29. Scandella, C. J., Kilpatrick, J., Lidster, W., Parker, C., Moore, J. P., Moore, G. K., Mann, K. A., Brown, P., Coates, S., Chapman, B. *et al.* (1993) *AIDS Res. Hum. Retroviruses* **9,** 1233–1244.
30. Kaiser, P. K., Offermann, J. T. & Lipton, S. A. (1990) *Neurology* **40,** 1757–1761.
31. Lipton, S. A., Sucher, N. J., Kaiser, P. K. & Dreyer, E. B. (1991) *Neuron* **7,** 111–118.
32. Lipton, S. A. (1991) *Ann. Neurol.* **30,** 110–114.
33. Lipton, S. A. (1994) *Nature (London)* **367,** 113–114.
34. Lei, S. Z., Pan, Z.-H., Aggarwal, S. K., Chen, H.-S. V., Hartman, J., Sucher, N. J. & Lipton, S. A. (1992) *Neuron* **8,** 1087–1099.
35. Lipton, S. A., Choi, Y.-B., Pan, Z.-H., Lei, S. Z., Chen, H.-S. V., Sucher, N. J., Loscalzo, J., Singel, D. J. & Stamler, J. S. (1993) *Nature (London)* **364,** 626–632.
36. Chao, C. C., Hu, S., Sheng, W. S., Bu, D., Bukrinsky, M. I. & Peterson, P. K. (1996) *Glia* **16,** 276–284.
37. Bonini, J. A. & Steiner, D. F. (1997) *DNA Cell. Biol.* **16,** 1023–1030.
38. Luster, A. D. (1998) *N. Engl. J. Med.* **338,** 436–445.
39. Spleiss, O., Gourmala, N., Boddeke, H. W., Sauter, A., Fiebich, B. L., Berger, M. & Gebicke-Haerter, P. J. (1998) *J. Neurosci. Res.* **53,** 16–28.
40. Davis, C. B., Dikic, I., Unutmaz, D., Hill, C. M., Arthos, J., Siani, M. A., Thompson, D. A., Schlessinger, J. & Littman, D. R. (1997) *J. Exp. Med.* **286,** 1793–1798.
41. Cheng-Mayer, C., Liu, R., Landau, N. R. & Stamatatos, L. (1997) *J. Virol.* **71,** 1657–1661.
42. Ohagen, A., Ghosh, S., He, J., Huang, K., Chen, Y., Yuan, M., Osathanondh, R., Gartner, S., Shi, B., Shaw, G. & Gabuzda, D. (1999) *J. Virol.* **73,** 897–906.
43. Plesoff, O., Sol, N., Labrosse, B. & Alizon, M. (1997) *J. Virol.* **71,** 3259–3262.
44. Parolin, C. B. A., Choe, H., Farzan, M., Kolchinsky, P., Heesen, M., Ma, Q., Gerard, C., Palu, G., Dorf, M. E., Springer, T. & Sodroski, J. (1998) *J. Virol.* **72,** 1652–1656.
45. Najjar, V. A. (1979) *Adv. Exp. Med. Biol.* **121,** 131–147.
46. Bhat, N. R., Zhang, P., Lee, J. C. & Hogan, E. L. (1998) *J. Neurosci.* **18,** 1633–1641.
47. Kawasaki, H., Morooka, T., Shimohama, T., Gotoh, Y. & Nishida, E. (1997) *J. Biol. Chem.* **272,** 18518–18521.
48. Mukherjee, P. K., DeCoster, M. A., Campbell, F. Z., Davis, R. J. & Bazan, N. G. (1999) *J. Biol. Chem.* **274,** 6493–6498.

Journal of Neurochemistry, 2001, **79**, 1–8

REVIEW

Exploitation of the HIV-1 coat glycoprotein, gp120, in neurodegenerative studies *in vivo*

M. Tiziana Corasaniti,*[,1] Mauro Maccarrone,† Robert Nistico,‡ Walter Malorni,§ Domenicantonio Rotiroti* and Giacinto Bagetta‡

**Faculty of Pharmacy, University of Catanzaro 'Magna Graecia' and IBAF-CNR, Catanzaro, Italy*

†Department of Experimental Medicine and Biochemical Sciences, University of Rome 'Tor Vergata', Rome

‡Department of Pharmacobiology, Calabria University at Cosenza, and Experimental Neurobiology Center 'Mondino-Tor Vergata-Santa Lucia', University of Rome 'Tor Vergata', Rome, Italy

§Department of Ultrastructures, Istituto Superiore di Sanità, Rome, Italy

Abstract

Neuronal loss has often been described at post-mortem in the brain neocortex of patients suffering from AIDS. Neuroinvasive strains of HIV infect macrophages, microglial cells and multinucleated giant cells, but not neurones. Processing of the virus by cells of the myelomonocytic lineage yields viral products that, in conjunction with potentially neurotoxic molecules generated by the host, might initiate a complex network of events which lead neurones to death. In particular, the HIV-1 coat glycoprotein, gp120, has been proposed as a likely aetiologic agent of the described neuronal loss because it causes death of neurones in culture. More recently, it has been shown that brain neocortical cell death is caused in rat by intracerebroventricular injection of a recombinant gp120 coat protein, and that this occurs via apoptosis. The latter observation broadens our knowledge in the pathophysiology of the reported neuronal cell loss and opens a new lane of experimental research for the development of novel therapeutic strategies to limit damage to the brain of patients suffering from HIV-associated dementia.

Keywords: apoptosis, cyclooxygenase type-2 (COX-2), HIV-associated dementia (HAD), HIV-1 gp120, interleukin-1β, neocortex.

J. Neurochem. (2001) **79**, 1–8.

Cognitive impairment, postural disorders and tremor are among the most common symptoms encountered in patients suffering from HIV-associated dementia (HAD), a neurological syndrome described in some 20% of AIDS patients (Everall *et al.* 1994). The neuropathological features of the brain described at post-mortem are myelin pallor, appearance of multinucleated giant cells, infiltration by blood-derived macrophages, astroglial cell reaction and brain cortical neuronal cell loss (Everall *et al.* 1991; Price and Perry 1994). The syndrome has been attributed to infection of the brain caused by the human immunodeficiency virus type 1 (HIV-1) because it is observed in patients free from opportunistic infections or concomitant cancer in the brain (Price and Perry 1994), although neuroinvasive strains of HIV infect macrophages, microglial cells and multinucleated giant cells, but not neurones (Mucke *et al.* 1995). Processing of the virus by cells of the myelomonocytic lineage yields host and viral products known to initiate a complex network of events, which may lead neurones to death and to the development of cerebral atrophy in AIDS patients (Gray *et al.* 2000). In particular, the HIV-1 coat protein, gp120, has been proposed as a likely aetiologic agent of the described neuronal loss because it causes the death of neurones in culture (Lipton and Gendelman 1995). More recently, brain cortical cell death has also been reported following intracerebroventricular injection of gp120 in rat, and this occurs via apoptosis.

Received November 14, 2000; revised manuscript received May 9, 2001; accepted May 19, 2001.

Address correspondence and reprint requests to Prof. Giacinto Bagetta, Department of Pharmacobiology, University of Calabria at Cosenza, 87030 Arcavacata di Rende, Italy. E-mail: gbagett@tin.it

[1]Holds the Chair of Pharmacology.

Abbreviations used: COX-2, cyclooxygenase type-2; HAD, HIV-associated dementia; HIV-1, human immunodeficiency virus type-1; GFAP, glial fibrillary acidic protein; ICE, interleukin-converting enzyme; i.c.v., lateral cerebral ventricle; IL-1β, interleukin-1β; NGF, nerve growth factor; NO, nitric oxide; PGE_2, prostaglandin E_2.

Here, we summarize *in vivo* data supporting a role for the HIV coat protein, gp120, in the mechanisms of neuronal cell loss often described in the brain cortex of patients suffering from HAD (Everall *et al.* 1994).

gp120 causes apoptosis in the neocortex of rat

Transgenic animals overexpressing gp120 in astrocytes display a pattern of neuropathological changes reminiscent of those described in subjects with AIDS, thus supporting a role for the HIV-1 coat protein in the pathophysiology of the associated neurological syndrome (Toggas *et al.* 1994). Retardation in behavioural development has been described in neonatal rats treated systemically with gp120 (Glowa *et al.* 1992; Hill *et al.* 1993), demonstrating that this is capable of causing cognitive impairment along with neuronal damage. More recently, Barks *et al.* (1997) have reported that in P7 neonatal rats focal injection of gp120 into the CA1 area of one dorsal hippocampus failed to produce, five days later (P12), hippocampal atrophy, and also failed to cause neuronal damage other than a subtle focal pyramidal cell loss immediately adjacent to the injection track. In these animals, however, the same authors have shown that focal intrahippocampal co-injection of gp120 and NMDA brought the reduction of hippocampal volume caused by the latter excitotoxin from 19% to 26.4%; this effect was prevented by antagonists of the NMDA-receptor complex, thus providing direct evidence of neurotoxic synergism between the HIV-1 coat glycoprotein gp120 and excitatory amino acids *in vivo* in the immature brain, and confirming that this interaction may occur at the level of the NMDA subtype of glutamate receptor (Barks *et al.* 1997). Lack of gross hippocampal damage and of statistically significant neuronal cell loss has been previously reported in adult rats receiving focal injection of gp120 (Bagetta *et al.* 1994a,b), and this is in line with the data reported by Barks *et al.* (1997). More recently, using the terminal-transferase (terminal fluorescein 12-dUTP nick-end labeling, TUNEL) technique (Gavrieli *et al.* 1992), we have shown the occurrence of DNA fragmentation in brain cortical tissue sections of adult rats receiving injections of the viral protein into one lateral cerebral ventricle (i.c.v.; Bagetta *et al.* 1995, 1996a), suggesting that neuronal death caused by the HIV-1 coat protein may be of the apoptotic type. The latter deduction has been confirmed by transmission electron microscopy (TEM) analysis of brain tissue sections obtained from rats treated with gp120 that revealed compaction and marginalization of nuclear chromatin along the inner surface of the nuclear envelope, and convolution of the nuclear margin in brain cortical cells (Bagetta *et al.* 1996b) (Fig. 1), unequivocal signs of early and late apoptosis (Kerr *et al.* 1987). In these animals, ultrastructural changes indicative of late apoptosis, such as masses of condensed chromatin and clumping of the nuclear envelope, have also been seen along with enlargement of the endoplasmic reticulum and normally

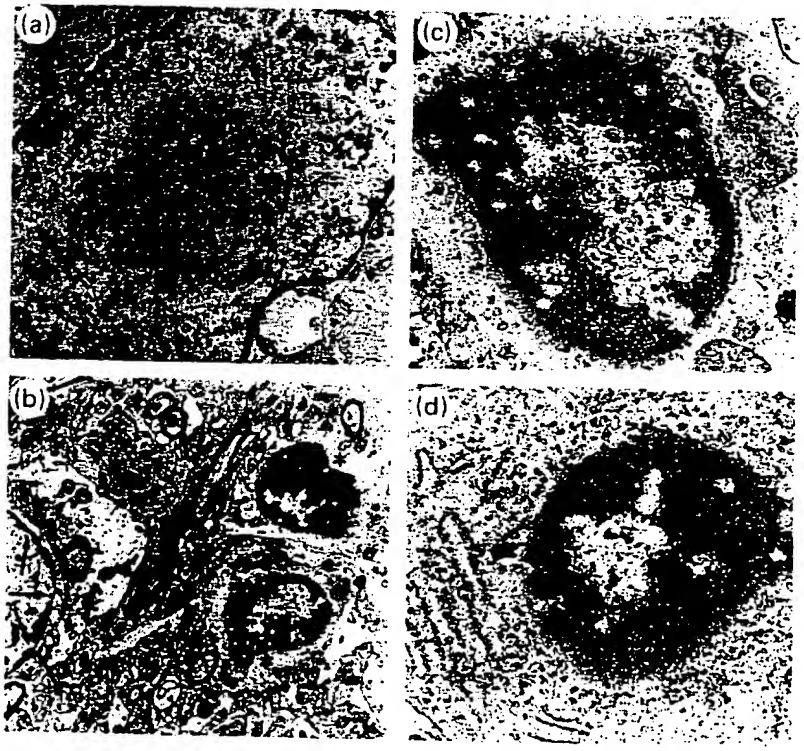


Fig. 1 (a) Nucleus of a neurone from a control, BSA-treated rat (100 ng given i.c.v. once daily for seven consecutive days) with normally dispersed chromatin (×6880). (b–d) Microphotographs showing apoptotic nuclei from the cortex of a rat receiving a single daily injection of gp120 (100 ng/day) for seven consecutive days. At low magnification (×540, b), two apoptotic nuclei (asterisks) and an injured cell with dilated nuclear envelope (arrows) can be seen. Chromatin aggregation, and pore dilation and clustering, typical of apoptotic cell death, are easely detectable at high magnification (×17 000). Note the change in mitochondrial integrity in (c) and endosplamic reticulum dilation in (d). Taken from Bagetta *et al.* (1996b).

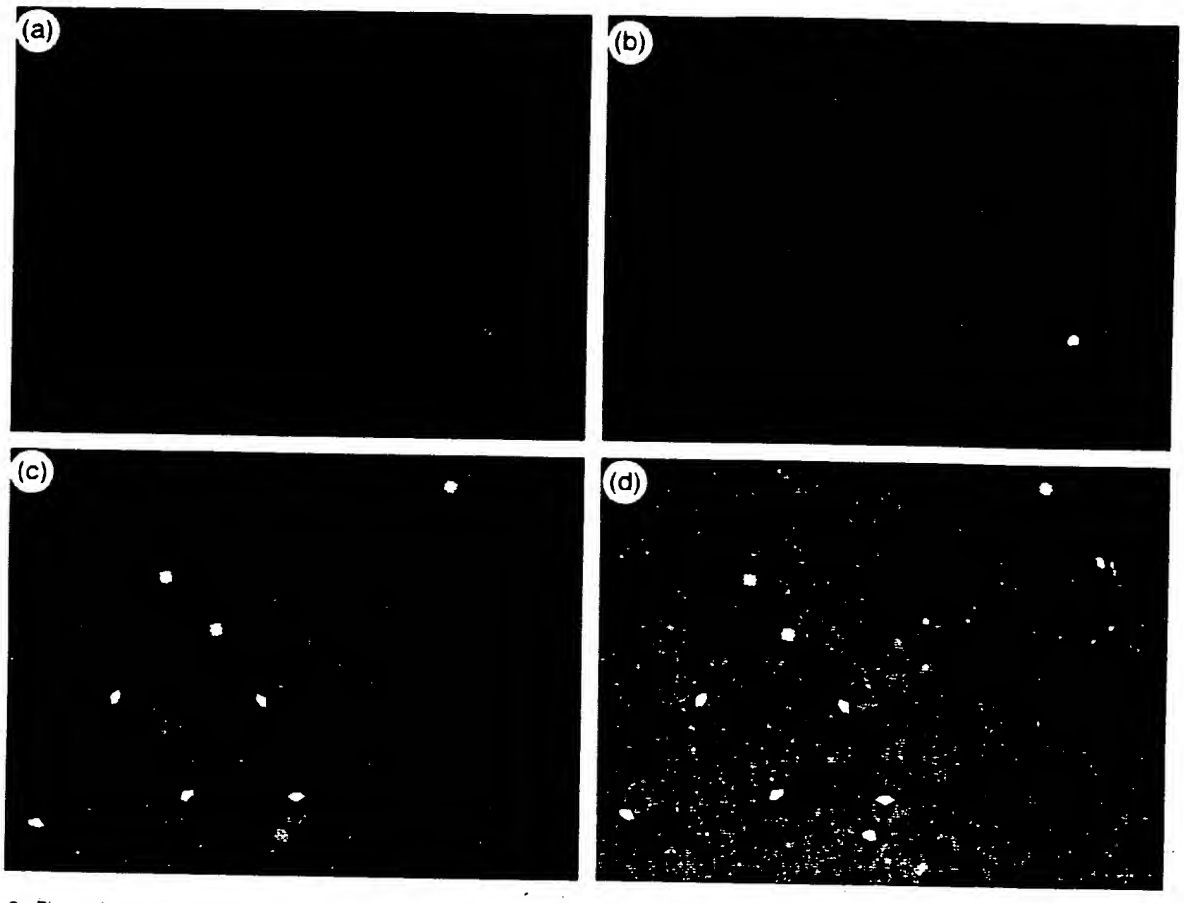


Fig. 2 Photomicrographs to show neurofilament-triplet (NF-T) and IL-1β double-immunoreactivity in brain tissue coronal sections (30 μm) obtained from a rat killed 24 h after a single i.c.v. injection of gp120 (100 ng), and processed for immunohistofluorescence. NF-T-immunopositivity (red fluorescence) is evident throughout panel (b); the specificity of NF-T immunostaining is confirmed by the lack of immunoreactivity in an adjacent section (a), incubated in the absence of the primary antibody for negative control. Green fluorescence, indicating specific IL-1β immunoreactivity (see panel c for negative control; the same section was shown in panel a), is shown in panel (d; the same section was shown in panel b). Arrowheads in (d) indicate cells double-immunopositive (see yellow dots) for IL-1β and NF-T (see panel b for comparison). Asterisks indicate cells positive for NF-T (b) and negative for IL-1β. Green immunofluorescence is also evident (panel d) in areas of the tissue section lacking cell bodies, and this may conceivably represent secretory IL-1β. Magnification: 40×. Reprinted from Bagetta *et al.* (1999) with permission from Elsevier Science.

appearing mitochondria. Immunoelectronmicroscopy analysis of brain neocortical cells bearing ultrastructural features typical of apoptosis revealed that these are immunopositive for the neurofilament (Fig. 2), a typical neuronal marker (Bagetta *et al.* 1999); by contrast, glial fibrillary acidic protein (GFAP) immunopositive cells appeared normal, suggesting that under the present experimental conditions astroglial cells may not undergo apoptosis (data not shown; Bagetta *et al.* 1999).

Neuronal apoptosis by gp120 was minimized in rats receiving (1 h beforehand) a single daily injection (0.25 pmoles given i.c.v. for seven consecutive days in all instances) of the β-chemokines RANTES, MIP-1α (natural ligands for the CCR5 chemokine receptor) or the α-chemokine SDF-1α (natural ligand for CXCR4 chemokine receptor) (Meucci and Miller 1999); likewise gp120, a higher dose (2.5 pmoles) of SDF-1α caused *in situ* DNA fragmentation (Corasaniti *et al.* 2000a). Collectively, these data support the concept that neuronal and microglial mechanisms, downstream CCR5 and CXCR4 receptors, coreceptors for gp120 binding and HIV-1 penetration into macrophages and T cells, respectively (Meucci and Miller 1999), may be responsible for neuronal apoptosis caused by the HIV-1 coat protein in the neocortex of rat.

gp120 causes abnormal xpression of int rleukin-1β in the neocort x

The mechanisms through which gp120 causes apoptosis in the brain of rat has yet to be discovered, although a series

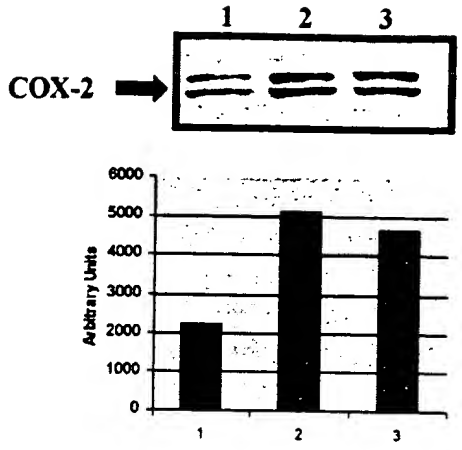


Fig. 3 A single i.c.v. dose of the HIV-1 coat protein gp120 causes a rapid enhancement of COX-2 expression in the brain neocortex of rat. This is a typical example of western blotting analysis to show COX-2 expression in neocortical brain tissue homogenate obtained from two independent rats treated 6 h previously with a single dose of bovine serum albumin (BSA, 100 ng/i.c.v.; lane 1) or gp120 (100 ng/i.c.v.; lane 2), respectively. Lane 3 shows the effect of MK801 (0.3 mg/kg given i.p. 30 min before gp120) on gp120-induced enhanced expression of COX-2 (lane 2). For comparison, the histograms show relative intensity values of the autoradiographic bands (see above) as determined by computer-assisted densitometric analysis (QUANTISCAN, Biosoft, Cambridge, UK). Note that gp120 almost doubles the expression of COX-2 as compared with control (BSA treated), and this is unaffected by treatment with MK801, a selective antagonist of the NMDA subtype of glutamate receptor. Taken from Corasaniti *et al.* (2000a).

of recent experimental data does implicate the pro-inflammatory cytokine interleukin-1β (IL-1β). In fact, immunohistochemical and western blotting experiments show that treatment with gp120 enhances the expression of IL-1β in the neocortex, and double-labelling immunofluorescence experiments have established that neuronal and, possibly, microglial cells are the main source of IL-1β (Bagetta *et al.* 1999) (Fig. 3). Immunoelectron microscopy and enzyme-linked immunosorbent assay (ELISA) data have established that IL-1β is expressed, although at very low levels, in the mitochondria of brain neocortical cells of naive untreated rats; more importantly, subchronic administration (i.c.v.) of gp120 enhances the mitochondrial expression of the pro-inflammatory cytokine, and this implicates *in situ* activation of interleukin-converting enzyme (ICE) (Corasaniti *et al.* 2001). In agreement with the latter deduction, antagonism studies have shown that combined treatment with gp120 and the inhibitor II (Ac-Tyr-Val-Ala-Asp-chloromethylketone) of ICE (Milligan *et al.* 1995), the protease (also known as caspase 1) that processes pro-IL-1β in biologically mature IL-1β (Black *et al.* 1988; Kostura *et al.* 1989; Yuan *et al.* 1993; Walker *et al.* 1994; Martins and Earnshow 1997), minimizes apoptotic cell death induced by the viral protein in the neocortex of rat (Bagetta *et al.* 1999). Quite importantly, treatment with the antagonist of IL-1 receptor (IL-1ra), the receptor species that mediates most of the biological actions of IL-1β (Dripps *et al.* 1991; Hagan *et al.* 1996), prevents apoptotic cell death caused by the viral protein (Corasaniti *et al.* 1998; Bagetta *et al.* 1999) and, likewise for gp120, subchronic i.c.v. administration of murine recombinant IL-1β causes apoptosis in the neocortex of rat (Bagetta *et al.* 1999), further implicating this cytokine in the mechanism of gp120-induced neocortical cell death.

Cyclooxygenase-2 induction by gp120 triggers apoptosis via an excitotoxic, glutamate mediated, mechanism

The mechanism through which IL-1β mediates gp120-induced apoptosis in the neocortex of rat is not known. In the mammalian brain this pro-inflammatory cytokine represents a physiological signal for secretion of nerve growth factor (NGF), and this could enhance the survival of injured neurones (Strijbos and Rothwell 1995). Interestingly, i.c.v. injections of gp120 enhanced IL-1β expression (Bagetta *et al.* 1999) but failed to elevate NGF production in the neocortex (Bagetta *et al.* 1996a), and this might contribute, at least in part, to cell death (Bagetta *et al.* 1995, 1996b) because of the lack of adequate trophic support (see Corasaniti *et al.* 1998 for further discussion).

It is well established that IL-1β can also affect the expression of inducible enzymes, such as nitric oxide synthase (iNOS) and cyclooxygenase-2 (COX-2), the terminal products of which may be highly cytotoxic (Merrill *et al.* 1993). However, at variance with several *in vitro* data (Lipton and Gendelman 1995), in rats treated with gp120 we failed to observe significant changes in brain cortical citrulline (Bagetta *et al.* 1996a, 1997, 1998a), the coproduct of NO synthesis (Knowles and Moncada 1994). Although these data do negate the occurrence of excessive NO production in the neocortex of gp120-treated rats, it cannot be excluded that physiological levels of NO can interact with other radical species that may originate from activated brain cortical microglial cells (Bagetta *et al.* 1999) to produce peroxynitrite, known to spontaneously decompose to yield the hydroxyl radical, a species even more cytotoxic than NO and known to be involved in apoptosis (Coyle and Puttfarcken 1993).

Instead, more recent data do support an important role for COX-2 in the mechanism of gp120-induced apoptosis. In fact, we have reported immunohistochemical evidence demonstrating that subchronic i.c.v. treatment with gp120 enhances the expression of COX-2 in the neocortex of rat (Bagetta *et al.* 1998b). More importantly, a single dose of gp120 causes an increase of COX-2 expression, which is

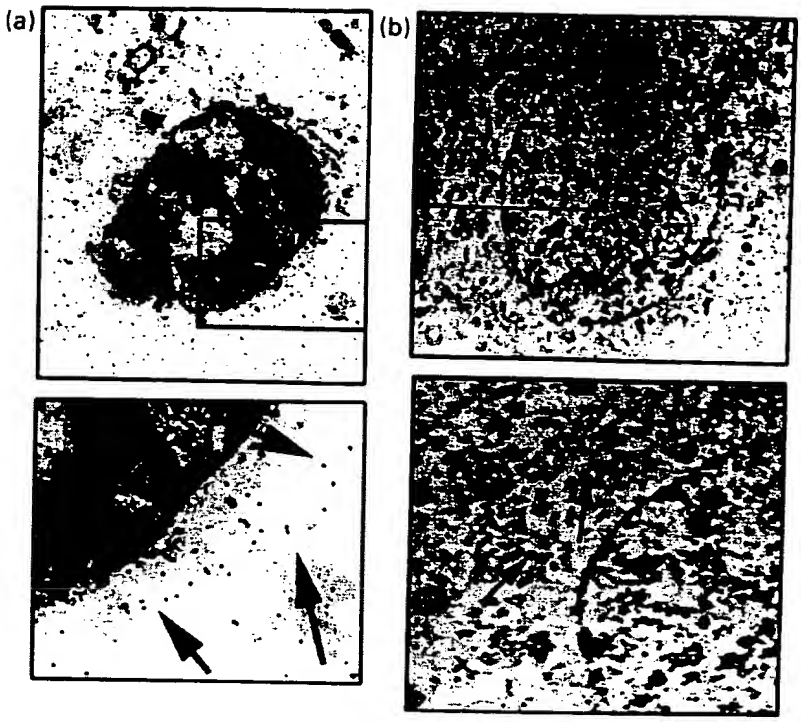


Fig. 4 Ultrastructural evidence of apoptosis caused by the HIV-1 coat glycoprotein gp120 in a rat brain neocortical cell immunopositive for the neurofilament, cytoskeletal, proteins and typical neuronal markers. Ultrathin tissue sections from the brain neocortex of a rat treated for seven consecutive days with a single daily dose (100 ng, given i.c.v.) of gp120 have been processed for electron microscopy detection of neurofilament protein immunogold positive cells (see Bagetta *et al.* 1999 for antisera properties). In (a), the upper panel (11 000× magnification) shows masses of condensed chromatin typical of late apoptosis; the lower panel shows, at a higher magnification (33 000×) of the area indicated by the box in the upper panel, a typical filament distribution (see arrows) of alligned gold particles. In (b), the lower panel shows rare gold particles (see arrows; 13 000× magnification). These are indistinguishable in the boxed area at lower magnification (6000×) of the upper panel in a neocortical brain tissue section adjacent to the one shown in (a) and processed for negative control.

apparent 6 h after the injection of the viral protein (Fig. 4), and this is paralleled by a significant accumulation of prostaglandin E_2 (PGE_2) in the neocortex (Corasaniti *et al.* 2000b; Maccarrone *et al.* 2000) and a significant increase in body temperature (Bagetta *et al.* 1999) in the rat. Experimental evidence suggests that in the mammalian CNS, enhanced expression of COX-2, and accumulation of products of the arachidonic acid cascade, including tromboxan B_2 and PGE_2, may be implicated in the pathophysiology of brain damage that follows exposure to

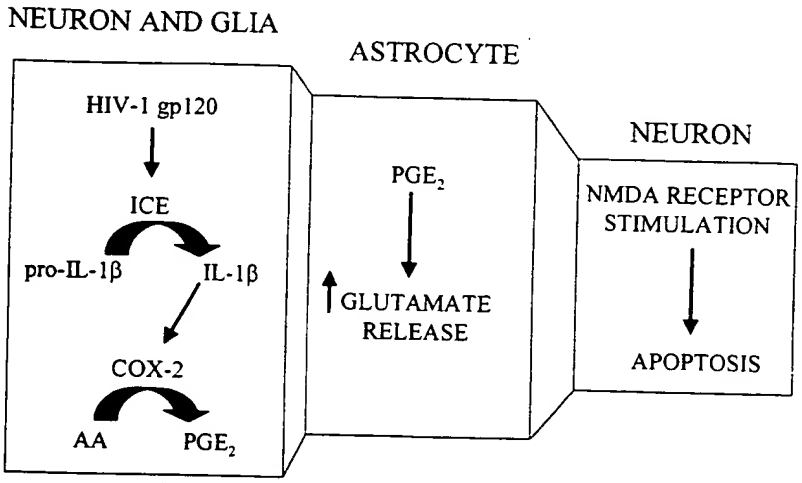


Fig. 5 Schematic representation of a unifying hypothesis on the mechanisms underlying neuronal apoptosis induced by gp120 in the neocortex of rat. Administration of recombinant HIV-1 gp120 IIIB enhances neuronal and microglial expression of IL-1β (Bagetta *et al.* 1999), an event that requires the conversion of the pro-IL-1β in the mature form of this cytokine via the intervention of interleukin converting enzyme (ICE) (Corasaniti *et al.* 2001). IL-1β may enhance the expression of COX-2 (it is established that in the mammalian brain COX-2 is located in neuronal cells; Yamagata *et al.* 1993) to convert arachidonic acid (AA) into prostaglandin E_2 (PGE_2) which than accumulate (Bagetta *et al.* 1998b; Maccarrone *et al.* 2000). Elevated PGE_2 stimulate Ca^{2+}-dependent release of glutamate from astrocytes (Bezzi *et al.* 1998) and this may be responsible for excitotoxic neuronal apoptosis in the neocortex of rat (Corasaniti *et al.* 2000a).

excitotoxic stimuli (Gaudet *et al.* 1980; Fostermann *et al.* 1982; Baran *et al.* 1987; Seregi *et al.* 1987; Planas *et al.* 1995; Nogawa *et al.* 1997). Therefore, it is conceivable that the observed abnormal expression of COX-2 and that accumulation of PGE_2 may be implicated in the mechanisms of apoptosis caused by gp120 in the neocortex of rat. In agreement with the latter hypothesis is the observation that apoptosis induced by gp120 is reduced by a systemic pretreatment with indomethacin (Bagetta *et al.* 1998b), a specific but non-selective inhibitor of COX activities, and by NS398, a selective COX-2 inhibitor (Corasaniti *et al.* 2000a).

Under physiologic conditions, the level of expression of COX-2 gene product appears to correlate well with the state of activation of excitatory, glutamate-mediated, synaptic transmission (Yamagata *et al.* 1993). *In vitro* and *in vivo* data suggest that gp120 enhances glutamate transmission via the release from astroglial cells of not yet well-identified excitotoxins acting at the NMDA, but not non-NMDA, receptors in the mammalian brain (Lipton and Gendelman 1995). Altogether, these data support the concept that the enhanced expression of COX-2 and the accumulation of PGE_2 observed here may be the consequence of abnormal activation of glutamate neurotransmission in the neocortex of gp120-treated rat. However, this does not appear to be the case because under our experimental conditions a systemic pretreatment with MK801, a selective antagonist of the NMDA receptor complex, failed to counteract gp120-enhanced COX-2 expression observed 6 h after treatment with the viral coat protein (Corasaniti *et al.* 2000a) (Fig. 4). However, systemic pretreatment with competitive and non-competitive NMDA receptor antagonists or with U-74389G, a free radical scavenger of the 21-aminosteroid family, reduced gp120-induced apoptosis in the neocortex of rat (Corasaniti *et al.* 2000b), supporting an excitotoxic glutamate-mediated mechanism of death (Choi 1988). Bezzi *et al.* (1998) have previously demonstrated that products of the arachidonic acid cascade (PGE_2 being among the most potent) stimulate the Ca^{2+}-dependent release of glutamate from astroglial cells, leading to the suggestion that this mechanism may have physiological as well as pathophysiological consequences in the mammalian brain. Therefore, to rationalize the observed lack of MK801 effect on COX-2 expression with the neuroprotection afforded by the NMDA receptor antagonists and by the 21-aminosteroid, U-74389G, we suggest that IL-1β may be responsible for the gp120-evoked rapid induction of COX-2 and accumulation of PGE_2, which may elevate, possibly through a mechanism similar to that described by Bezzi *et al.* (1998), synaptic glutamate; this would then trigger a vitious loop leading the cell to oxidative stress and apototic death via an excitotoxic mechanism (Choi 1988). The series of events initiated by gp120 and leading to apoptotic cell death are schematically reported in Fig. 5.

In conclusion, the observation that gp120 induces apoptotic cell death in the rat neocortex *in vivo*, together with the recent evidence of DNA fragmentation reported at post-mortem in the brain of AIDS patients (Petito and Roberts 1995), suggests that this mechanism may underlie the well-established cortical neuronal loss described in the brain of AIDS patients. The recent immunolocalization of gp120 in human brain tissue with the neuropathological correlates of HIV-1 encephalitis and pre-mortem diagnosis of HAD provides the missing link in the understanding of HIV neuropathogenisis (Jones *et al.* 2000); gp120 may, in fact, be present in sufficient quantity during HIV infection to cause neuronal damage (Jones *et al.* 2000), although other viral components, such as Tat, may also contribute (Bansal *et al.* 2000).

Here, we would like to speculate that confirmation of the neuroinflammatory steps we have partly dissected in the brain of gp120-treated rats may prove useful for the study of the underlying pathophysiological mechanisms of neuronal death. Finally, demonstration at the ultrastructural level of the occurrence of apoptosis in the brain cortex of AIDS patients will validate the usefulness of the rat model we have developed for the characterization of the neuroprotective profile of drugs that interfere with mediators of neuroinflammation and the crucial steps involved in the activation of the death programme.

Acknowledgements

This work was funded by: the II AIDS Project to AFA and WM, Istituto Superiore di Sanità, Rome, Italy; the MURST-DAE Cofin '1998' (D.M. 1451/97); the European Fund for Regional Development (Calabria Region, P.O.P. 94/99). Our thanks go to Mr N. Fico and to Mr G. Politi for excellent technical assistance.

References

Bagetta G., Corasaniti M. T., Finazzi-Agro' A. and Nistico' G. (1994a) Does HIV-1 glycoprotein GP120 produce neurodegeneration? *Trends Pharmacol. Sci.* **15**, 362–363.

Bagetta G., Palma E., Finazzi-Agrò A. and Nisticò G. (1994b) Intracerebral injection of human immunodeficiency virus type 1 coat glycoprotein GP120 does not produce neurodegeneration in rat. *Neurosci. Lett.* **176**, 97–100.

Bagetta G., Corasaniti M. T., Berliocchi L., Navarra M., Finazzi-Agrò A. and Nisticò G. (1995) HIV-1 gp120 produces DNA fragmentation in the cerebral cortex of rat. *Biochem. Biophys. Res. Commun.* **211**, 130–136.

Bagetta G., Corasaniti M. T., Aloe L., Berliocchi L., Costa N., Finazzi-Agrò A. and Nisticò G. (1996a) Intracerebral injection of human immunodeficiency virus type 1 coat protein gp120 differentially affects the expression of nerve growth factor and nitric oxide synthase in the hippocampus of rat. *Proc. Natl Acad. Sci. USA* **93**, 928–933.

Bagetta G., Corasaniti M. T., Malorni W., Rainaldi G., Costa N., Berliocchi L., Finazzi-Agrò A. and Nisticò G. (1996b) The HIV-1 gp120 causes ultrastructural changes typical of apoptosis in the rat cerebral cortex. *Neuroreport* **7**, 1722–1724.

Bagetta G., Corasaniti M. T., Costa N., Berliocchi L., Finazzi-Agrò A. and Nisticò G. (1997) The human immunodeficiency virus type 1 (HIV-1) glycoprotein gp120 reduces the expression of neuronal nitric oxide synthase in the hippocampus but not in the cerebral cortex and medial septal nucleus. *Neurosci. Lett.* **224**, 1–4.

Bagetta G., Berliocchi L., Costa N., Palma E., Aloe L. and Nisticò G. (1998a) Role of nitric oxide and NGF in the mechanisms of neurotoxicity induced by the HIV-1 coat protein gp120 in rat, in *Nitric Oxide and the Cell: Proliferation, Differentiation and Death* (Moncada S., Nisticò G., Bagetta G. and Higgs A. E., eds), pp. 145–155. Portland Press, London.

Bagetta G., Corasaniti M. T., Paoletti A. M., Berliocchi L., Nisticò R., Giammarioli A. M., Malorni W. and Finazzi-Agrò A. (1998b) HIV-1 gp120-induced apoptosis in the rat neocortex involves enhanced expression of cyclo-oxygenase type 2 (COX-2). *Biochem. Biophys. Res. Commun.* **244**, 819–824.

Bagetta G., Corasaniti M. T., Berliocchi L., Nisticò R., Giammarioli A. M., Malorni W., Aloe L. and Finazzi-Agrò A. (1999) Involvement of interleukin-1β in the mechanism of human immunodeficiency virus type 1 (HIV-1) recombinant protein gp120-induced apoptosis in the neocortex of rat. *Neuroscience* **89**, 1051–1066.

Bansal A. K., Mactutus C. F., Nath A., Maragos W., Hauser K. F. and Booze R. M. (2000) Neurotoxicity of HIV-1 proteins gp120 and Tat in the rat striatum. *Brain Res.* **879**, 42–49.

Baran H., Heldt R. and Hertting G. (1987) Increased prostaglandin formation in rat brain following systemic application of kainic acid. *Brain Res.* **404**, 107–112.

Barks J. D. E., Liu X.-H., Sun R. and Silverstein F. S. (1997) gp120, human immunodeficiency virus-1 coat protein, augments excitotoxic hippocampal injury in perinatal rats. *Neuroscience* **76**, 397–409.

Bezzi P., Carmignoto G., Pasti L., Vesce S., Rossi D., Rizzini B. L., Pozzan T. and Volterra A. (1998) Prostaglandins stimulate calcium-dependent glutamate release in astrocytes. *Nature* **391**, 281–285.

Black R. A., Kronheim S. R., Cantrell M., Deeley M. C., March C. J., Prickett K. S., Wignall J., Conlon P. J., Cosman D., Hopp T. P. and Mochizuki D. Y. (1988) Generation of biologically active interleukin-1 of IL-1β by proteolytic cleavage of the inactive precursor. *J. Biol. Chem.* **263**, 9437–9442.

Choi D. W. (1988) Glutamate neurotoxicity and diseases of the nervous system. *Neuron* **1**, 623–634.

Corasaniti M. T., Bagetta G., Rotiroti D. and Nistico′ G. (1998) The HIV envelope protein, gp120, in the nervous system. Interactions with nitric oxide, IL-1β and NGF signalling, with pathological implications *in vivo* and *in vitro*. *Biochem. Pharmacol.* **56**, 153–156.

Corasaniti M. T., Strongoli M. C., Piccirilli S., Nisticò R., Costa A., Bilotta A., Turano P., Finazzi-Agro′ A. and Bagetta G. (2000a) Apoptosis induced by gp120 in the neocortex of rat involves enhanced expression of cyclooxygenase type 2 (COX-2) and is prevented by NMDA receptor antagonists and by the 21-aminosteroid U-74389G. *Biochem. Biophys. Res. Commun.* **274**, 664–669.

Corasaniti M. T., Strongoli M. C., Piccirilli S., Nistico′ R., Finazzi-Agro′ A. and Bagetta G. (2000b) Evidence that CCR5 and CXCR4 chemokine receptors are involved in the mechanisms of HIV-1 gp120-induced apoptosis in the neocortex of rat. *Br. J. Pharmacol.* **129**, 114P.

Corasaniti M. T., Turano P., Bilotta A., Malorni W., Stringaro A. R., Nisticò R., Finazzi-Agro A. and Bagetta G. (2001) Evidence that increases of mitochondrial immunoreactive IL-1β by HIV-1 gp120 implicate *In Situ* cleavage of pro-IL-1β in the neocortex of rat. *J. Neurochem.* **78**, 611–618.

Coyle J. T. and Puttfarcken P. (1993) Oxidative stress, glutamate, and neurodegenerative disorders. *Nature* **262**, 689–695.

Dripps D. J., Brandhuber B. C., Thompson R. C. and Eisenberg S. P. (1991) Interleukin-1 receptor antagonist binds to the 80-kDa interleukin-1 receptor, but does not initiate interleukin-1 signal transduction. *J. Biol. Chem.* **266**, 10331–10336.

Everall I. P., Luthert P. J. and Lantos P. (1991) Neuronal loss in the frontal cortex in HIV infection. *Lancet* **337**, 1119–1121.

Everall I. P., Glass J. D., McArthur J., Spargo E. and Lantos P. (1994) Neuronal density in the superior frontal and temporal gyri does not correlate with the degree of human immunodeficiency virus-associated dementia. *Acta Neurpathol.* **88**, 538–544.

Fostermann U., Seregi A. and Hertting G. (1982) Anticonvulsive effects of endogenous prostaglandins formed in brain of spontaneously convulsing gerbils. *Prostaglandins* **27**, 913–923.

Gaudet R. J., Alan I. and Levin L. (1980) Accumulation of cyclooxygenase products of arachidonic acid metabolism in gerbil brain during reperfusion after bilateral common carotid artery occlusion. *J. Neurochem.* **35**, 653–658.

Gavrieli Y., Sherman Y. and Ben-Sasson S. A. (1992) Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J. Cell. Biol.* **119**, 493–501.

Glowa J. R., Panlilio L. V., Brenneman D. E., Gozes I., Fridkin M. and Hill J. M. (1992) Learning impairment following intracerebral administration of the HIV envelope protein gp120 or VIP antagonist. *Brain. Res.* **603**, 222–233.

Gray F., Adle-Biassette H., Brion F., le Maner I., Levy V. and Corcket G. (2000) Neuronal apoptosis in human immnunodeficiency virus infection. *J. Neurovirol.* **6**, S38–S43.

Hagan P., Poole S., Bristow A. F., Tilders F. and Silverstein F. S. (1996) Intracerebral NMDA injection stimulates production of interleukin-1β in perinatal rat brain. *J. Neurochem.* **67**, 2215–2218.

Hill J. M., Mervis R. F., Avidor R., Moody T. W. and Brenneman D. E. (1993) HIV envelope protein-induced neuronal damage and retardation of behavioural development in rat neonates. *Brain Res.* **603**, 222–233.

Jones M. V., Bell J. E. and Nath A. (2000) Immunolocalization of HIV envelope gp120 in HIV encephalitis with dementia. *AIDS* **14**, 2709–2713.

Kerr J. F. R., Searle J., Harmon B. V. and Bishop C. J. (1987). *Perspecitves on Mammalian Cell Death*. (Potten C. S., ed.), pp. 93–128. Oxford University Press, Oxford.

Knowles R. G. and Moncada S. (1994) Nitric oxide synthases in mammals. *Biochem. J.* **298**, 249–258.

Kostura M. J., Tocci M. J., Limjuci G., Chin J., Cameron P., Hillman A. G., Chartrain N. A. and Schmidt J. A. (1989) Identification of a monocyte specific pre-interleukin 1 of IL-1β convertase activity. *Proc. Natl Acad. Sci. USA* **86**, 5227–5231.

Lipton S. A. and Gendelman H. E. (1995) Dementia associated with the acquired immunodeficiency syndrome. *New Engl. J. Med.* **332**, 934–940.

Maccarrone M., Bari M., Corasaniti M. T., Nistico′ R., Bagetta G. and Finazzi-Agrò A. (2000) HIV-1 coat glycoprotein gp120 induces apoptosis in rat brain neocortex by deranging the arachidonate cascade in favour of prostanoids. *J. Neurochem.* **75**, 196–203.

Martins L. M. and Earnshow W. C. (1997) Apoptosis: alive and kicking in 1997. *Trends Cell Biol.* **7**, 111–114.

Merrill J. E., Ignarro L. J., Sherman M. P., Melinek J. and Lane T. E. (1993) Microglial cell cytotoxicity of oligodendrocytes is mediated through nitric oxide. *J. Immunol.* **151**, 2132–2141.

Meucci O. and Miller O. (1999) AIDS and the brain: is there a chemokine connection? *Trends Neurosci.* **22**, 471–479.

Milligan C. E., Prevette D., Yaginuma H., Homma S., Cardwell C., Fritz L. C., Tomaselli K. J., Oppenheim R. W. and Schwartz L. M. (1995) Peptide inhibitors of the ICE family arrest programmed cell death of motoneurons in vivo and in vitro. *Neuron* **15**, 385–393.

Mucke L., Masliah E. and Campbell I. L. (1995) Transgenic models to assess the neuropathogenic potential of HIV-1 proteins and cytokines. *Curr. Top. Microbiol. Immunol.* **202**, 187–205.

Nogawa S., Zhang F., Ross M. E. and Iadecola C. (1997) Cyclooxygenase-2 gene expression in neurons contributes to ischemic brain damage. *J. Neurosci.* **17**, 2746–2755.

Petito C. K. and Roberts B. (1995) Evidence of apoptotic cell death in HIV encephalitis. *Am. J. Pathol.* **146**, 1121–1130.

Planas A. M., Soriano M. A., Rodriguez-Ferràe E. and Ferrer I. (1995) Induction of cyclooxygenase-2 mRNA and protein following transient focal ischemia in the rat brain. *Neurosci. Lett.* **200**, 187–190.

Price R. W. and Perry S. W. (1994). *HIV, AIDS, and the Brain*. Raven Press, New York.

Seregi A., Keller M. and Hertting G. (1987) Are cerebral prostanoids of astroglial origin? *Brain Res.* **404**, 113–120.

Strijbos P. J. and Rothwell N. J. (1995) Interleukin-1beta attenuates excitatory amino acid induced neurodegeneration in vitro: involvement of nerve growth factor. *J. Neurosci.* **15**, 3468–3474.

Toggas S. M., Masliah E., Rockenstein E. M., Rall G. F., Abraham C. R. and Mucke L. (1994) Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice. *Nature* **367**, 188–193.

Walker N. P. C., Talanian R. V., Brady K. D., *et al.* (1994) Crystal structure of the cysteine protease interleukin-1 of IL-1β-converting enzyme: a $(p20/p10)_2$ homodimer. *Cell* **78**, 343–352.

Yamagata K., Andreasson K. I., Kaufmann W. E., Barnes C. A. and Worley P. F. (1993) Expression of a mitogen-inducible cyclooxygenase in brain neurones: regulation by synaptic activity and glucocorticoids. *Neuron* **11**, 371–386.

Yuan J., Shaham S., Ledoux S., Ellis H. M. and Horvitz H. R. (1993) The C. elegans cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1β-converting enzyme. *Cell* **75**, 641–652.

Annu. Rev. Med. 2003. 54:285–303
doi: 10.1146/annurev.med.54.101601.152143

AIDS-RELATED MALIGNANCIES

David T. Scadden
AIDS Research Center and Cancer Center, Massachusetts General Hospital, Harvard Medical School, 149 13th Street, Room 5212, Boston, Massachussetts 02129; e-mail: scadden.david@mgh.harvard.edu

Key Words Kaposi's sarcoma, non-Hodgkin's lymphoma, Epstein-Barr virus, human herpesvirus-8, human immunodeficiency virus

■ **Abstract** Immunodeficiency alters the risk of cancer. Specific types of immune dysfunction are associated with different tumor risks, but most tumors are related to oncogenic viruses. In acquired immunodeficiency due to the human immunodeficiency virus (HIV), HIV itself rarely directly causes cancer; rather, it provides the immunologic background against which other viruses can escape immune control and induce tumors. The most common malignancies are Kaposi's sarcoma and non-Hodgkin's lymphoma. This chapter discusses the pathophysiologic background of these tumors, how they have been affected by the use of anti-HIV medications, and their clinical management.

INTRODUCTION

The number of individuals infected with the human immunodeficiency virus type 1 (HIV-1) worldwide is currently estimated at 40 million. Manifestations are highly dependent on geographic location, genetic background, and most importantly the availability of antiretroviral therapy. Malignancy, a complication of HIV-induced and other forms of immunodeficiency, is restricted to a limited spectrum of tumors (Table 1), generally those for which an infectious cofactor has been defined (Table 2). Although the mechanisms by which these tumors arise vary markedly with tumor type and virus, inadequate immunologic control provides a unifying conceptual framework among them. As opportunistic malignancies, these tumors have the potential for responsiveness to immunologic control and the development of novel therapeutics. The two major types of tumors seen in the setting of HIV are Kaposi's sarcoma and non-Hodgkin's lymphoma. These are the focus of this chapter.

Epidemiology

The spectrum of tumors in the context of HIV-1 infection varies according to risk group and has been substantially influenced by the advent of combination, highly active antiretroviral therapy (HAART). Kaposi's sarcoma (KS) is the tumor

0066-4219/03/0218-0285$14.00

TABLE 1 Tumor types with increased incidence in HIV disease

Definite	Possible
Kaposi's sarcoma	Seminoma
Non-Hodgkin's lymphoma	
Squamous cell neoplasia	
Hodgkin's disease	
Leiomyosarcoma (in children)	
Plasmacytoma	

most obviously affected by this potent treatment. Although the incidence of this infection-related neoplasm was already declining in the United States prior to the availability of HIV protease inhibitor therapy, the therapy has made it a relative rarity among treated HIV-infected individuals (1–3). Both regression of KS following successful HIV suppression on HAART and a marked decrease in KS incidence since the availability of HAART have been noted, with estimates of decline as high as 80-fold (2–6). In settings where HAART is not available, such as sub-Saharan Africa, KS remains a major problem and is the major cancer diagnosis in some regions (6a,b).

Like KS, primary central nervous system lymphomas (PCNS)—a subset of non-Hodgkin's lymphomas—have undergone dramatic changes in incidence. Although this complication of far-advanced HIV disease was much less common than KS, and therefore its decline is less well documented, U.S. centers that had previously seen cases monthly are now seeing them annually. PCNS is an agonal manifestation of AIDS, and like post-transplant lymphoproliferative disease, it is virtually uniformly associated with the presence of Epstein-Barr virus (EBV)

TABLE 2 Secondary virus infections associated with AIDS-related malignancies

Virus	Tumor
Epstein-Barr virus (EBV)	Non-Hodgkin's lymphoma (PCNS, some systematic ARLs, oropharyngeal T-cell)*
	Hodgkin's disease
	Leiomyosarcoma (children)
Kaposi's sarcoma herpesvirus (KSHV)	Kaposi's sarcoma
	Non-Hodgkin's lymphoma (primary effusion lymphoma)
Human papillomavirus (HPV)	Squamous cell neoplasia

*Abbreviations: ARLs, AIDS-related lymphomas; PCNS, primary central nervous system.

in the tumor tissue. Comparable to post-transplant lymphoproliferative disease, the profile of EBV latent gene expression includes EBNA1-6 and LMP-1, -2, a type III pattern seen when EBV is used to transform B cells in vitro (7, 8). Among these gene products are those readily targeted by cytotoxic T lymphocytes, which may account for the marked reduction in incidence of PCNS among patients with successful control of HIV-induced immune destruction by HAART. Of note, even among those with detectable EBV-specific cytotoxic T lymphocytes, abnormalities in cell function have been noted and associated with EBV lymphoproliferation (9).

In contrast to the PCNS subset of AIDS-related lymphomas (ARLs), the risk of systemic lymphomas is less dramatically reduced by HAART (3, 10). Overall, the estimated decline in systemic lymphomas is approximately two- to sevenfold since the introduction of HAART (1, 11–13). The largest study to date was an observational cohort analysis of 8500 HIV-positive individuals across Europe (EuroSIDA) (12). The incidence of all subtypes of lymphoma was significantly reduced after 1999, when the use of HAART was commonplace, compared with the period prior to HAART (marked in this study as beginning in September 1995). Similarly, an international multicohort study found a reduction of approximately twofold following the introduction of HAART (11). Of note, this series assessed subtypes of lymphomas and observed the greatest difference in immunoblastic lymphoma and PCNS. Burkitt's lymphoma and Hodgkin's disease appeared to be largely unaffected (11). The changes evident within only some lymphoma subsets does suggest possible differential involvement of immune function in tumor development.

NON-HODGKIN'S LYMPHOMA

Pathophysiology

There are ARLs, such as PCNS, in which EBV is uniformly present and for which a pathophysiologic process may be readily envisioned. In that setting, EBV latent genes are expressed in a type III pattern including expression of the latent membrane protein–2 (LMP-2), which is known to dysregulate cell growth control and can transform B lymphocytes. The systemic lymphomas appear to have a more complex pathophysiology, however. EBV is present in a subset of these tumors (33%–67% depending on the report), and the type III latent gene pattern is not consistently observed (14–16). Some of these tumors appear to express a profile of EBV genes more consistent with Hodgkin's disease, and the large proportion of those without EBV have a range of other genetic abnormalities. Among AIDS-related large-cell lymphomas, Bcl-6 rearrangement, c-myc rearrangement, and p53 mutations occur in approximately 33%, 40%, and 25%, respectively (17). The small-cell (Burkitt's and Burkitt's-like) histology subset is commonly associated with c-myc rearrangements, but not Bcl-6 and rarely p53 mutations (18–21). There is no clear link between EBV and any specific genetic mutation other than those noted for the histologic subtype (18–21). Among

those tumors in which c-myc is rearranged, c-myc is transposed into the immunoglobulin gene heavy-chain switch region (20, 22–25), which strongly suggests that the rearrangement occurred at the time of class switching rather than during early B cell differentiation. Because this follows VDJ recombination of the immunoglobulin locus, the cell of origin is likely to be a post-germinal-center B cell.

B cell growth kinetics appear to be altered in the presence of HIV infection and clinically manifest as the frequent lymphadenopathy and hypergammaglobulinemia seen in this group of patients. HIV may directly contribute to the process through antigenic drive, and there are reports that HIV envelope glycoprotein may directly enhance B cell activation (26, 27). HIV gp120 envelopes capable of interacting with the CXCR-4 chemokine receptor, in particular, may effect changes in B cell proliferation, as this receptor is known to provide a growth-promoting signal to B cell subsets (28–32). Perturbation of the T cell compartment, with enhancement of TH2 subpopulations and release of B cell stimulatory interleukins, IL-10 and IL-4, probably further augments proliferation (33, 34). With control of HIV replication, the B cell stimulus may be reduced through these direct and indirect mechanisms, resulting in a decrease in hypergammaglobulinemia with successful HAART.

Genetic analyses of patient cohorts have begun to reveal host-related factors relevant to the risk of lymphoma. Individuals with polymorphisms in regulatory regions of the chemokine gene encoding stroma-derived growth factor–1 (SDF-1) were noted to have an excess risk of developing lymphoma, particularly of the Burkitt's subtype (35). Although the specific mechanism has not been shown, SDF-1 is the cognate ligand for CXCR-4, is a known B cell growth factor, and may provide an excessive proliferative stimulus. HIV-infected individuals heterozygous for an inactivating deletion mutation of CCR5 (CCR5Δ32) were noted to have a threefold decrease in lymphoma risk (36). This abnormality may decrease the sensitivity of target cells to the chemokine RANTES, which may result in altered B cell function, either directly or through T cell–mediated events (36). Further genomic analysis of the host-pathogen interaction is clearly an area of potential for defining patients with variable risk and may ultimately lead to screening or preventative strategies.

Clinical Presentation, Evaluation, and Treatment

Systemic ARLs frequently involve tissues outside of lymph nodes and therefore have a wide array of possible clinical presentations. Common extranodal sites include the gastrointestinal tract, bone marrow, and central nervous system (CNS), though virtually any tissue may be involved (19, 38–54). Histologic subsets do have some discriminating patterns of involvement. For example, large-cell tumors preferentially involve the gastrointestinal tract and small-cell tumors the bone marrow and meninges (45, 55). The presenting symptoms of lymphoma do not appear to be appreciably affected by HAART (56, 57).

Owing to a high incidence of CNS involvement noted early in the HIV epidemic [20% in one study (43)], it has become commonplace to more aggressively evaluate the CNS in patients with systemic ARL. This has generally included imaging and cerebrospinal fluid sampling studies, and many centers prophylactically administer intrathecal therapy to all patients. Particular attention should be paid to those in whom EBV is documented in the primary tumor, since in one study its presence strongly predicted an increased risk for CNS relapse ($p = 0.003$) (58). The same study also defined extranodal involvement as a strong predictive factor ($p = 0.006$). Whether such criteria can be used to subselect patients in whom CNS prophylaxis may be restricted has not been tested formally, but the data do support targeting intrathecal chemotherapy to those with EBV in the tumor tissue and those with extranodal involvement of high-risk sites such as marrow, testis, or paranasal sinus (59).

The prognosis for patients with ARL prior to HAART was poor but appears to be changing with the overall improvement in health and tolerance of chemotherapy afforded by control of HIV. Most prognostic factors were defined before HAART and may need to be revised to accommodate broader, more current experience. However, the largest multivariant analysis to date indicated that CD4 count <100 cells/mm^3, age >35 years, intravenous drug use, and stage III/IV disease were negative prognostic factors (60). When one or none of these factors was present, the overall survival was 46 weeks; with two factors, 44 weeks; with three or four factors, 18 weeks.

The International Prognostic Index (IPI) (61) is a useful means of stratifying risk in aggressive lymphomas outside the context of AIDS but has not been broadly applied to date in ARL. A study of 46 patients did indicate that high IPI score was predictive of poor outcome (62), and other reports have indicated that factors used in the IPI such as elevated LDH (63) or age >40 years provide independent prognostic information in ARL. In the context of HAART, it is likely that IPI can be used to define risk in ARL and will be tested in current trials. Burkitt's or Burkitt's-like histology has not been consistently noted to be of prognostic significance. Treatment protocols to date have generally included this subset of patients with other histologic groups and not detected a distinct outcome. However, as more information is gained in the era of HAART, now that other HIV complications contribute less to outcome, this histologic subset may distinguish itself as more problematic. Whether more aggressive treatment programs should be applied to this group in the setting of HIV disease remains undecided.

Primary effusion lymphoma is a rare form of systemic lymphoma associated with AIDS. It is a liquid-phase hematologic malignancy that rarely involves the blood or lymph nodes and generally does not present with a tumor mass. Rather, a body cavity effusion (64–66) laden with large anaplastic or immunoblastic-appearing cells is the hallmark. The cells immunophenotypically mark with surface CD45 (common leukocyte antigen) but do not stain with antibodies specific for B cell (CD20 or CD19) or T cell (CD3) antigens. Molecular analysis of tumor cells does demonstrate VDJ rearrangement of the immunoglobulin locus, confirming a

B cell origin. Unique among the ARLs, primary effusion lymphoma cells also are uniformly found to contain the Kaposi's sarcoma herpesvirus (KSHV) genome and frequently demonstrate coinfection with EBV. These tumors are not restricted to HIV-related immunodeficiency and may be found in other immunodeficient states. They provide a unique and intriguing paradigm for virus-induced human malignancy.

Therapy

The impact of HAART on lymphoma risk has been paralleled by improved treatment tolerance in patients with lymphoma. The ability of patients to receive full-dose therapy has now been well established, and the options of intensive dosing and transplantation are being explored. Prior to the availability of HAART, the limited prognosis and poor tolerance of therapy pushed experimentation to pursue minimally toxic regimens. A phase III randomized trial comparing full-dose with half-dose m-BACOD demonstrated equivalent tumor outcomes with a more favorable toxicity profile for the lower-dose regimen (67). This study set a standard for reduced-dose approaches, which has now been supplanted as HAART has improved the overall health of the patients. Low-dose regimens are now generally reserved for those with advanced AIDS who have either failed HAART or for whom HAART is not available. Studies that have not formally compared dose intensity, but in which different dose levels were used, have indicated a more favorable effect on tumor outcomes with standard-dose regimens (68). Therefore, CHOP and its equivalents have resumed their position as the up-front treatment of choice for patients with ARL.

Studies with modified dosing schedules indicate that infusional regimens may benefit ARL patients. The CDE regimen of Sparano and colleagues has yielded response rates of ~58% (69), and a study by the U.S. National Cancer Institute using dose-adjusted EPOCH (70, 71) demonstrated durable responses in >75% of patients (72). These are the most encouraging data to date and if validated may set a new standard for this patient group. Whether adding rituxan to standard chemotherapy conveys benefit in the setting of ARL is not clear. A randomized, phase III trial comparing CHOP alone with CHOP plus rituxan has recently been completed by the U.S. National Cancer Institute AIDS Malignancy Consortium and should provide important new information.

Given the improved tolerance of therapy with HAART, transplantation has again been considered for patients with ARL. This approach, which proved highly toxic and showed very poor results early in the HIV epidemic (73–81), now appears to be far more promising. Small studies in the United States and Europe have indicated that autologous transplant is well tolerated, with no delay in engraftment or undue opportunistic complications (82, 83). Furthermore, long-term survival in the context of relapsed Hodgkin's or non-Hodgkin's lymphoma and HIV have been reported (84).

Genetic manipulation of stem cells to render them resistant to HIV has been a conceptually appealing but thus far disappointing strategy (83, 85). Allogeneic

or minimally myeloablative approaches are now entering clinical trial. Such approaches can only be recommended in the context of clinical trials at present, given the complex interplay of immune function, viral replication, and tumor biology in these patients.

Although prior HAART can increase tolerance of antitumor medication, there is controversy as to whether HAART can be given concurrently with antitumor medication. In an effort to resolve this issue, the AIDS Malignancy Consortium studied stavudine, lamivudine, and indinivir at fixed dose in combination with CHOP chemotherapy. No untoward or unexpected toxicities were observed. The pharmacokinetics of doxorubicin and indinivir were unaffected, but a ~50% reduction in cyclophosphamide clearance was observed without apparent clinical impact (86). Although these data indicate the relative safety of concurrent HAART and antitumor medication, they are restricted to a small subset of antiretroviral drugs and a regimen less common now than at the time of the study. Recognizing the potential complexity of regimens and potential drug-drug interactions, the National Cancer Institute stopped all antiretrovirals during its trial of modified EPOCH chemotherapy (87). As anticipated, the HIV viral load increased and CD4 cell count decreased, but both parameters normalized following reintroduction of HAART at the end of antitumor therapy. Transiently discontinuing antiretrovirals during cancer chemotherapy had no apparent deleterious effects. However, this is a heavily weighted emotional issue for many patients, and thoughtful discussion with each individual is necessary when considering whether to stop anti-HIV medications.

KAPOSI'S SARCOMA

Viral Epidemiology

Kaposi's sarcoma (KS) is the most common neoplasm associated with AIDS, but not all HIV-infected individuals are at risk for it. It is more common in geographic regions associated with endemic KS, such as the Mediterranean basin and sub-Saharan Africa, and is particularly likely to occur in patients who acquired HIV by male homosexual activity. The disproportionate risk for KS among select immunodeficient populations raised the suspicion of a secondary infectious factor, which was confirmed by the identification of KSHV (88, 88a). Comparative genetic analysis of KS-involved tissue with normal tissue revealed DNA homologous with viral sequences from the gammaherpesvirus family. This group contains at least two other viruses capable of transforming human cells: EBV, which immortalizes human B cells, and *Herpesvirus saimiri*, which immortalizes human T cells (88). KSHV is a 165-kb, double-stranded DNA virus (89) that is present in patients prior to tumor formation (89, 90), has a high seroprevalence in populations with a high incidence of KS (91), and is present in cells composing the tumors (89). These data provide compelling evidence for a causative association of KSHV with KS.

Definitive seroepidemiologic studies of KSHV infection await broadly accepted assays, but data from a number of approaches have begun to outline the rates

of infection in some populations. The ORF73 gene product is the serodominant antigen. Assays for it have high specificity, but their sensitivity is only ~80% in HIV-infected populations with clinical KS (91). The prevalence of KSHV in the United States as determined by this assay has been reported to be 1%–2% of blood donors, 2% of hemophiliacs, 3%–4% of HIV-positive women (92), and 25%–30% of HIV-positive homosexual men (93). A whole-virus lysate assay provides greater sensitivity (92% positivity among patients with KS) and detected 11% positivity among healthy blood donors (94). Thus, in prevalence, KSHV resembles *Herpes simplex* rather than the virtually ubiquitous EBV, at least among North Americans and northern Europeans. The epidemiology is quite different in sub-Saharan Africa and the Mediterranean basin, where prevalence rates exceed 40% in some populations.

How KSHV is transmitted remains unclear. That male homosexual activity is associated with transmission is quite clear from a longitudinal study of men in San Francisco followed over a ten-year period. That study demonstrated that KSHV seroconversion risk was linearly related to the number of male-male sexual intercourse contacts (93). Men who had in excess of 250 sexual partners in the preceding two years had a seropositivity rate of 65%. Other modes of transmission must occur, based on the epidemiology of the disease, but are less well defined. In Africa, childhood infection occurs after the risk of vertical transmission but prior to sexual activity. KSHV has been documented in saliva and oral transmission has some epidemiologic support (95), although the spread of the virus by oral contamination is thought to be inefficient.

Pathology and Pathogenesis

KSHV infection is necessary but not sufficient for KS. Its malignant potential appears to be quite low outside the setting of immune compromise, but it is present in sporadic endemic and epidemic settings of KS. KSHV enters cells by engaging a cellular integrin receptor ($\alpha 3/\beta 1$, CD49c/29) (96). It can infect a range of different cell types, including B cells and dermal microvascular endothelial cells (97). It is present in KS tissues but is rapidly lost from culture when KS-derived cells are propagated in culture (97). The basis for the KSHV induction of tumor remains controversial and may be distinct from the paradigms proposed for other viral-related tumors. Although it is in the same herpesvirus subfamily as EBV, the latent genes implicated in EBV-induced transformation do not have homologues in KSHV. *Herpesvirus saimiri* encodes a transforming gene product that does have homology to a KSHV gene, *K1*, and that gene product has transforming ability when transfected into target cells (98). However, *K1* is expressed in the lytic and not the latent phase of the KSHV life cycle. Other KSHV gene products have been associated with transformation in transfection assays, but their gene expression profile is not consistent with the concept that latent program genes are those likely to be involved in transformation. For example, both the KSHV gene *K9*, which encodes a homologue of the interferon regulatory factor family, and *K12*, which

has no clear gene family homology, can transform cells. Perhaps the most intriguing gene product is a constitutively activated chemokine receptor–like protein, ORF74, which can transform cells (103) and induce a disease closely resembling KS when expressed in mice (104). It may be, therefore, that lytic phase genes may contribute to oncogenesis in *trans*, influencing the function of neighboring cells while the lytically infected cell dies. Clinical data do provide some indirect support for this unconventional paradigm: Medications that affect lytic replication of herpesviruses, ganciclovir and foscarnet, have been associated with antitumor effects (105–108). Further analysis of how this virus affects tumor growth awaits definition of methods for propagating the virus.

The KSHV genome encodes a number of gene products, which have the potential for affecting cells in *trans*. There are two homologues for chemokine genes, vMIP-I (K6) and vMIP-II (K4), and a viral IL-6 homologue, K2. Each interacts with cell surface receptors with either agonist (K2 and K6) or antagonist (K4) effects (109, 110). The IL-6 homologue is a particularly attractive candidate for influencing normal cell proliferation, but circulating levels of vIL-6 do not correlate with tumor development (111).

Host response to the virus appears to be critical in determining the clinical outcome of infection, including tumor development. The association of KS with immunodeficiency is clear, and evidence for complete regression of tumor with either HAART or, in the setting of organ transplant, reduced immunosuppressive medication further demonstrates the importance of immune control (112, 113). KSHV, like other members of the herpesvirus family, has evolved mechanisms to avoid immune attack. MHC class I cell surface expression is reduced by the viral gene products K5 and K3 because of enhanced endocytosis (114, 115) and reduced tapasin expression (116). It has also been demonstrated that K5 downregulates ICAM and B7-2, critical immune-modulating surface molecules for activation of effector cells (117). Therefore, host and viral mechanisms may dually contribute to inadequate immunologic control of KSHV.

It is not clear why HIV infection is particularly permissive of KS compared with other immunodeficiency states, but several mechanisms have been proposed. The HIV-1 tat gene product can enhance KSHV replication (118) and increase expression of IL-6 (119) and IL-6 receptor (120). HIV-1 replication may thereby directly potentiate KSHV effects and indirectly contribute to oncogenesis.

Treatment

The diagnosis of KS should not prompt a reflexive move to treat. This tumor may progress in an indolent manner even in patients with advanced immunosuppression. The decision to treat is based on the tumor's location, extent, and rapidity of change. For all patients, a critical aspect of tumor control is optimizing anti-HIV therapy. Response of pre-existing KS to HAART alone has been documented in up to 86% of patients (121), a rate exceeding that of most cytotoxic chemotherapy studies. These responses are generally durable and gradually

increase over time; in one multi-institutional study, only 6 of 39 KS patients treated with HAART still required KS-specific therapy 24 months after initiation of HAART (122). However, although HAART plays an important role, it is often insufficient in those with aggressive disease, and given the potential for aggressive or symptomatic KS to worsen prognosis (123), tumor-specific therapy may be indicated.

Tumor treatment may be locally applied for those with limited, accessible lesions and includes topical liquid nitrogen, intralesional vinblastine, and radiation therapy. In a randomized trial of 82 patients, topical 9-cis–retinoic acid cream demonstrated a sixfold higher response rate than placebo (124). However, local erythema and irritation were common effects that may offset the benefit in tumor control. For patients with edema, extensive mucocutaneous disease, or symptomatic pulmonary or gastrointestinal involvement, systemic chemotherapy is appropriate and generally well tolerated. Response rates in the literature are somewhat difficult to interpret because no standard measuring system has been applied, and the typical response criterion of changing bi-dimensional area may be misleading, since a residual hemosiderin stain is common even with histologic regression of KS. Single agents such as bleomycin and vincristine (125) or the combination of doxorubicin, bleomycin and vincristine (126) have demonstrated response rates of 57% to 88% (127). These drugs are often associated with toxicity, but this can be mitigated by more recent therapies of liposomal anthracyclines or paclitaxel. Because KS lesions are composed of vessels with poor integrity, liposomally encapsulated drugs are deposited in them. Drug concentrations have been found almost tenfold higher in lesions than in surrounding tissue (128). Two phase III studies, each involving ~250 HIV-positive KS patients, have evaluated liposomal doxorubicin. Superior tumor response (1.5–2-fold improvement) was observed relative to either bleomycin plus vincristine or that combination plus adriamycin (129, 130). A phase III study comparing liposomal daunorubicin with combined doxorubicin, bleomycin, and vincristine demonstrated a superior toxicity profile with no major difference in tumor response rates (131). No comparison of the liposomal agents has been reported. Despite the potential difference in tumor activity and minor differences in toxicity profile [for example, liposomal doxorubicin is associated with the hand-foot syndrome and liposomal daunorubicin is not (129)], the agents are often used interchangeably.

Paclitaxel, a tubulin stabilizer, has emerged as a highly active and generally very well-tolerated agent for KS. A phase I trial involving 28 patients demonstrated a major response in 71% (132), including individuals with heavily pretreated, anthracycline-treated KS. Low-dose paclitaxel (100 mg/m^2 every 2 weeks) is extremely well tolerated, and a phase II study reported a 59% response rate with a longer duration of response than was seen with other cytotoxic therapies for KS (133). Durability of the response to any cytotoxic agent is transient, and patients generally require chronic therapy unless anti-HIV therapy has permitted substantial immune regeneration. The cure for KS appears to be immune reconstitution, as cytotoxic agents are strictly palliative.

Antiangiogenic compounds are a natural strategy for combating this highly vascular tumor, and some trials have demonstrated encouraging results. Thalidomide is an angiogenesis inhibitor, and a phase II trial demonstrated a partial response in 4 of 13 patients over a 52-week period (134). The membrane metalloproteinase inhibitor, col-3, has been shown to be active in early-phase testing and is now entering phase II trial through the AIDS Malignancy Consortium. In contrast, Fumigillin (TNP-470) had little antitumor effect in a study of 38 patients (135), and IM862 has not demonstrated benefit in phase III testing. How the antiangiogenesis will be used, either alone or in combination, and how immunologic manipulation may ultimately contribute to the armamentarium against KS remain to be determined. However, the uniquely accessible and highly vascular nature of KS offers a particularly attractive target for testing angiogenic-modulating therapies.

CONCLUSION

The malignancies that complicate HIV disease represent a unique intersection of virology, immunology, and tumor biology. As such, they provide opportunities for furthering our understanding of cancer and for testing novel paradigms of therapy. Further study of these tumors offers insights that will reach beyond the HIV epidemic and may provide unique opportunities for evaluating new cancer treatment strategies.

The *Annual Review of Medicine* is online at http://med.annualreviews.org

LITERATURE CITED

1. Grulich AE, Li Y, McDonald AM, et al. 2001. Decreasing rates of Kaposi's sarcoma and non-Hodgkin's lymphoma in the era of potent combination antiretroviral therapy. *AIDS* 15:629–33
2. Rabkin CS, Testa MA, Huang J, Von Roenn JH. 1999. Kaposi's sarcoma and non-Hodgkin's lymphoma incidence trends in AIDS Clinical Trial Group study participants. *J. Acquir. Immune Defic. Syndr.* 21(Suppl. 1): S31–33
3. Biggar RJ. 2001. AIDS-related cancers in the era of highly active antiretroviral therapy. *Oncology (Huntingt.)* 15:439–48; discussion 48–49
4. Jacobson LP, Yamashita TE, Detels R, et al. 1999. Impact of potent antiretroviral therapy on the incidence of Kaposi's sarcoma and non-Hodgkin's lymphomas among HIV-1–infected individuals. Multicenter AIDS Cohort Study. *J. Acquir. Immune Defic. Syndr.* 21(Suppl. 1): S34–41
5. Buchbinder SP, Holmberg SD, Scheer S, et al. 1999. Combination antiretroviral therapy and incidence of AIDS-related malignancies. *J. Acquir. Immune Defic. Syndr.* 21(Suppl. 1):S23–26
6. Rabkin CS. 2000. AIDS and cancer in the era of highly active antiretroviral therapy (HAART). *Eur. J. Cancer* 37:1316–19

6a. Wabinga HR, Parkin DM, Wabwire-Mangen F, Mugerwa J. 1993. Cancer in Kampala, Uganda. Changes in incidence in the era of AIDS. *Int. J. Cancer* 54:5

6b. Chatlynne LG, Ablashi DV. 1999. Seroepidemiology of Kaposi's sarcoma-associated herpesvirus (KSHV). *Semin. Cancer Biol.* 3:175–85
7. Young L, Alfieri C, Hennessy K, et al. 1989. Expression of Epstein-Barr virus transformation-associated genes in tissues of patients with EBV lymphoproliferative disease. *N. Engl. J. Med.* 321:1080–85
8. Kieff E, Liebowitz D. 1996. Epstein-Barr virus and its replication. In *Virology*, ed. B Fields, D Knipe, p. 1889. Philadelphia: Lippincott-Raven
9. van Baarle D, Hovenkamp E, Callan MF, et al. 2001. Dysfunctional Epstein-Barr virus (EBV)–specific $CD8^+$ T lymphocytes and increased EBV load in HIV-1 infected individuals progressing to AIDS-related non-Hodgkin lymphoma. *Blood* 98:146–55
10. Grulich AE. 1999. AIDS-associated non-Hodgkin's lymphoma in the era of highly active antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* 21(Suppl. 1):S27–30
11. International Collaboration on HIV and Cancer. 2000. Highly active antiretroviral therapy and incidence of cancer in human immunodeficiency virus–infected adults. *J. Natl. Cancer Inst.* 92:1823–30
12. Kirk O, Pedersen C, Cozzi-Lepri A, et al. 2001. Non-Hodgkin lymphoma in HIV-infected patients in the era of highly active antiretroviral therapy. *Blood* 98:3406–12
13. Besson C, Goubar A, Gabarre J, et al. 2001. Changes in AIDS-related lymphoma since the era of highly active antiretroviral therapy. *Blood* 98:2339–44
14. Shibata D, Weiss LM, Hernandez AM, et al. 1993. Epstein-Barr virus-associated non-Hodgkin's lymphoma in patients infected with the human immunodeficiency virus. *Blood* 81(8):2102–9
15. Hamilton-Dutoit SJ, Pallesen G, Franzmann MB, et al. 1991. AIDS-related lymphoma. Histopathology, immunophenotype, and association with Epstein-Barr virus as demonstrated by in situ nucleic acid hybridization. *Am. J. Pathol.* 138:149–63
16. Hamilton-Dutoit SJ, Pallesen G, Karkov J, et al. 1989. Identification of EBV-DNA in tumour cells of AIDS-related lymphomas by in-situ hybridisation. *Lancet* 1:554–52
17. Gaidano G, Lo Coco F, Ye BH, et al. 1994. Rearrangements of the BCL-6 gene in acquired immunodeficiency syndrome–associated non-Hodgkin's lymphoma: association with diffuse large-cell subtype. *Blood* 84:397–402
18. Shiramizu B, Herndier B, Meeker T, et al. 1992. Molecular and immunophenotypic characterization of AIDS-associated, Epstein-Barr virus–negative, polyclonal lymphoma. *J. Clin. Oncol.* 10:383–89
19. Levine AM. 1992. Acquired immunodeficiency syndrome–related lymphoma. *Blood* 80:8
20. Pelicci PG, Knowles DM, Arlin ZA, et al. 1986. Multiple monoclonal B cell expansions and c-myc oncogene rearrangements in acquired immune deficiency syndrome–related lymphoproliferative disorders. Implications for lymphomagenesis. *J. Exp. Med.* 164:2049–60
21. Ballerini P, Gaidano G, Gong JZ, et al. 1993. Multiple genetic lesions in acquired immunodeficiency syndrome–related non-Hodgkin's lymphoma. *Blood* 81: 166–76
22. Chaganti RS, Jhanwar SC, Koziner B, et al. 1983. Specific translocations characterize Burkitt's-like lymphoma of homosexual men with the acquired immunodeficiency syndrome. *Blood* 61:1265–68
23. Petersen JM, Tubbs RR, Savage RA, et al. 1985. Small noncleaved B cell Burkitt-like lymphoma with chromosome t(8;14) translocation and Epstein-Barr virus nuclear-associated antigen in a homosexual man with acquired immune deficiency syndrome. *Am. J. Med.* 78:141–48
24. Haluska FG, Russo G, Kant J, et al. 1989. Molecular resemblance of an AIDS-associated lymphoma and endemic Burkitt

lymphomas: implications for their pathogenesis. *Proc. Natl. Acad. Sci. USA* 86: 8907–11

25. Neri A, Barriga F, Knowles DM, et al. 1988. Different regions of the immunoglobulin heavy-chain locus are involved in chromosomal translocations in distinct pathogenetic forms of Burkitt lymphoma. *Proc. Natl. Acad. Sci. USA* 85:2748–52
26. Kehrl JH, Rieckmann P, Kozlow E, Fauci AS. 1992. Lymphokine production by B cells from normal and HIV-infected individuals. *Ann. NY Acad. Sci.* 651:220–27
27. Yarchoan R, Redfield RR, Broder S. 1986. Mechanisms of B cell activation in patients with acquired immunodeficiency syndrome and related disorders. Contribution of antibody-producing B cells, of Epstein-Barr virus–infected B cells, and of immunoglobulin production induced by human T cell lymphotropic virus, type III/lymphadenopathy–associated virus. *J. Clin. Invest.* 78:439–47
28. Davis CB, Dikic I, Unutmaz D, et al. 1997. Signal transduction due to HIV-1 envelope interactions with chemokine receptors CXCR4 or CCR5. *J. Exp. Med.* 186:1793–98
29. Madani N, Kozak SL, Kavanaugh MP, Kabat D. 1998. gp120 envelope glycoproteins of human immunodeficiency viruses competitively antagonize signaling by coreceptors CXCR4 and CCR5. *Proc. Natl. Acad. Sci. USA* 95:8005–10
30. Popik W, Hesselgesser JE, Pitha PM. 1998. Binding of human immunodeficiency virus type 1 to CD4 and CXCR4 receptors differentially regulates expression of inflammatory genes and activates the MEK/ERK signaling pathway. *J. Virol.* 72:6406–13
31. Popik W, Pitha PM. 1998. Early activation of mitogen-activated protein kinase kinase, extracellular signal-regulated kinase, p38 mitogen-activated protein kinase, and c-Jun N-terminal kinase in response to binding of simian immunodeficiency virus to Jurkat T cells expressing CCR5 receptor. *Virology* 252:210–17
32. Su SB, Gong W, Grimm M, et al. 1999. Inhibition of tyrosine kinase activation blocks the down-regulation of CXC chemokine receptor 4 by HIV-1 gp 120 in $CD4^+$ T cells. *J. Immunol.* 162:7128–32
33. Clerici M, Wynn TA, Berzofsky JA, et al. 1994. Role of interleukin-10 in T helper cell dysfunction in asymptomatic individuals infected with the human immunodeficiency virus. *J. Clin. Invest.* 93:768–75
34. Muller F, Aukrust P, Nordoy I, Froland SS. 1998. Possible role of interleukin-10 (IL-10) and CD40 ligand expression in the pathogenesis of hypergammaglobulinemia in human immunodeficiency virus infection: modulation of IL-10 and Ig production after intravenous Ig infusion. *Blood* 92:3721–29
35. Rabkin CS, Yang Q, Goedert JJ, et al. 1999. Chemokine and chemokine receptor gene variants and risk of non-Hodgkin's lymphoma in human immunodeficiency virus-1–infected individuals. *Blood* 93:1838–42
36. Dean M, Jacobson LP, McFarlane G, et al. 1999. Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5-delta32 mutation. *Cancer Res.* 59:3561–64
37. Deleted in proof
38. Kalter SP, Riggs SA, Cabanillas F, et al. 1985. Aggressive non-Hodgkin's lymphomas in immunocompromised homosexual males. *Blood* 66:655–59
39. Ioachim HL, Dorsett B, Cronin W, et al. 1991. Acquired immunodeficiency syndrome–associated lymphomas: clinical, pathologic, immunologic, and viral characteristics of 111 cases. *Hum. Pathol.* 22:659–73
40. Ziegler JL, Beckstead JA, Volberding PA, et al. 1984. Non-Hodgkin's lymphoma in 90 homosexual men. Relation to generalized lymphadenopathy and the acquired immunodeficiency syndrome. *N. Engl. J. Med.* 311:565–70

41. Lowenthal DA, Straus DJ, Campbell SW, et al. 1988. AIDS-related lymphoid neoplasia. The Memorial Hospital experience. *Cancer* 61:2325–37
42. Deleted in proof
43. Levine AM, Wernz JC, Kaplan L. 1991. Low-dose chemotherapy with central nervous system prophylaxis and zidovudine maintenance in AIDS-related lymphoma. *JAMA* 266:84–88
44. Gill PS, Levine AM, Meyer PR, et al. 1985. Primary central nervous system lymphoma in homosexual men. Clinical, immunologic, and pathologic features. *Am. J. Med.* 78:742–48
45. Gill PS, Levine AM, Krailo M. 1987. AIDS-related malignant lymphoma: results of prospective treatment trials. *J. Clin. Oncol.* 5:1322–28
46. Knowles DM, Chamulak GA, Subar M. 1988. Lymphoid neoplasia associated with the acquired immunodeficiency syndrome (AIDS): the New York University Medical Center experience with 105 patients. *Ann. Intern. Med.* 108:744–53
47. Bermudez MA, Grant KM, Rodvien R, Mendes F. 1989. Non-Hodgkin's lymphoma in a population with or at risk for acquired immunodeficiency syndrome: indications for intensive chemotherapy. *Am. J. Med.* 86:71–76
48. Kaplan LD, Abrams DI, Feigal E. 1989. AIDS-associated non-Hodgkin's lymphoma in San Francisco. *JAMA* 261:719–24
49. Kaplan MH, Susin M, Pahwa SG, et al. 1987. Neoplastic complications of HTLV-III infection. Lymphomas and solid tumors. *Am. J. Med.* 82:389–96
50. Kaplan LD, Kahn JO, Crowe S, et al. 1991. Clinical and virologic effects of recombinant human granulocyte-macrophage colony–stimulating factor in patients receiving chemotherapy for human immunodeficiency virus–associated non-Hodgkin's lymphoma: results of a randomized trial. *J. Clin. Oncol.* 9:929–40
51. Remick SC, McSharry JJ, Wolf BC. 1993. Novel oral combination chemotherapy in the treatment of intermediate-grade and high-grade AIDS-related non-Hodgkin's lymphoma. *J. Clin. Oncol.* 11:1691–702
52. Freter CE. 1990. Acquired immunodeficiency syndrome–associated lymphomas. *J. Natl. Cancer Inst. Monogr.* 10:45–54
53. von Gunten CF, Von Roenn JH. 1992. Clinical aspects of human immunodeficiency virus–related lymphoma. *Curr. Opin. Oncol.* 4:894–99
54. Raphael M, Gentilhomme O, Tulliez M, et al. 1991. Histopathologic features of high-grade non-Hodgkin's lymphomas in acquired immunodeficiency syndrome. The French Study Group of Pathology for Human Immunodeficiency Virus–Associated Tumors. *Arch. Pathol. Lab. Med.* 115:15–20
55. Burkes RL, Meyer PR, Gill PS, et al. 1986. Rectal lymphoma in homosexual men. *Arch. Intern. Med.* 146:913–15
56. Levine AM, Seneviratne L, Espina BM, et al. 2000. Evolving characteristics of AIDS-related lymphoma. *Blood* 96: 4084–90
57. Matthews GV, Bower M, Mandalia S, et al. 2000. Changes in acquired immunodeficiency syndrome–related lymphoma since the introduction of highly active antiretroviral therapy. *Blood* 96:2730–34
58. Cingolani A, Gastaldi R, Fassone L, et al. 2000. Epstein-Barr virus infection is predictive of CNS involvement in systemic AIDS-related non-Hodgkin's lymphomas. *J. Clin. Oncol.* 18:3325–30
59. Scadden DT. 2000. Epstein-Barr virus, the CNS, and AIDS-related lymphomas: as close as flame to smoke. *J. Clin. Oncol.* 18:3323–24
60. Straus DJ, Huang J, Testa MA, et al. 1998. Prognostic factors in the treatment of human immunodeficiency virus–associated non-Hodgkin's lymphoma: analysis of AIDS Clinical Trials Group protocol 142—low-dose versus standard-dose m-BACOD plus granulocyte-macrophage

colony–stimulating factor. National Institute of Allergy and Infectious Diseases. *J. Clin. Oncol.* 16:3601–6

61. Shipp MA, Harrington DP, Klatt MM, et al. 1986. Identification of major prognostic subgroups of patients with large-cell lymphoma treated with m-BACOD or M-BACOD. *Ann. Intern. Med.* 104:757–65
62. Navarro JT, Ribera JM, Oriol A, et al. 1998. International prognostic index is the best prognostic factor for survival in patients with AIDS-related non-Hodgkin's lymphoma treated with CHOP. A multivariate study of 46 patients. *Haematologica* 83:508–13
63. Vaccher E, Tirelli U, Spina M, et al. 1996. Age and serum lactate dehydrogenase level are independent prognostic factors in human immunodeficiency virus–related non-Hodgkin's lymphomas: a single-institute study of 96 patients. *J. Clin. Oncol.* 14:2217–23
64. Cesarman E, Chang Y, Moore PS, et al. 1995. Kaposi's sarcoma–associated herpesvirus-like DNA sequences in AIDS-related body-cavity–based lymphomas. *N. Engl. J. Med.* 332:1186–91
65. Nador RG, Cesarman E, Chadburn A, et al. 1996. Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma–associated herpes virus. *Blood* 88:645–56
66. Karcher DS, Alkan S. 1997. Human herpesvirus-8–associated body cavity–based lymphoma in human immunodeficiency virus–infected patients: a unique B-cell neoplasm. *Hum. Pathol.* 28:801–8
67. Kaplan LD, Straus DJ, Testa MA, et al. 1997. Low-dose compared with standard-dose m-BACOD chemotherapy for non-Hodgkin's lymphoma associated with human immunodeficiency virus infection. *N. Engl. J. Med.* 336:1641–48
68. Ratner L, Lee J, Tang S, et al. 2001. Chemotherapy for human immunodeficiency virus–associated non-Hodgkin's lymphoma in combination with highly active antiretroviral therapy. *J. Clin. Oncol.* 19:2171–78
69. Sparano JA, Lee S, Chen M, et al. 1999. Phase II trial of infusional cyclophosphamide, doxorubicin, and etoposide (CDE) in HIV-associated non-Hodgkin's lymphoma: an Eastern Cooperative Oncology Group Trial. *Proc. Am. Soc. Clin. Oncol.* 18:12a (Abstr. 41)
70. Wilson WH, Bryant G, Bates S, et al. 1993. EPOCH chemotherapy: toxicity and efficacy in relapsed and refractory non-Hodgkin's lymphoma. *J. Clin. Oncol.* 11:1573–82
71. Gutierrez M, Chabner BA, Pearson D, et al. 2000. Role of a doxorubicin-containing regimen in relapsed and resistant lymphomas: an 8-year follow-up study of EPOCH. *J. Clin. Oncol.* 18:3633–42
72. Little R, Pearson D, Steinberg S, et al. 1999. *Dose-adjusted EPOCH chemotherapy (CT) in previously untreated HIV-associated non-Hodgkin's lymphoma (HIV-NHL).* Presented at Annu. Meet. Am. Soc. Clin. Oncology, 35th, May 15–18, Atlanta, GA
73. Holland HK, Saral R, Rossi JJ, et al. 1989. Allogeneic bone marrow transplantation, zidovudine, and human immunodeficiency virus type 1 (HIV-1) infection. Studies in a patient with non-Hodgkin lymphoma. *Ann. Intern. Med.* 111:973–81
74. Cooper MH, Maraninchi D, Gastaut JA, et al. 1993. HIV infection in autologous and allogeneic bone marrow transplant patients: a retrospective analysis of the Marseille bone marrow transplant population. *J. Acquir. Immune Defic. Syndr.* 6:277–84
75. Vilmer E, Rhodes-Feuillette A, Rabian C, et al. 1987. Clinical and immunological restoration in patients with AIDS after marrow transplantation, using lymphocyte transfusions from the marrow donor. *Transplantation* 44:25–29

76. Hassett JM, Zaroulis CG, Greenberg ML, Liegal FP. 1983. Bone marrow transplantation in AIDS. *N. Engl. J. Med.* 309:665
77. Bowden RA, Coombs RW, Nikora BH, et al. 1990. Progression of human immunodeficiency virus type-1 infection after allogeneic marrow transplantation. *Am. J. Med.* 88:49N–52N
78. Bardini G, Re MC, Rosti G, Belardinelli AR. 1991. HIV infection and bone-marrow transplantation. *Lancet* 337: 1163–64
79. Giri N, Vowels MR, Ziegler JB. 1992. Failure of allogeneic bone marrow transplantation to benefit HIV infection. *J. Paediatr. Child Health* 28:331–33
80. Torlontano G, DiBartolomeo P, DiGirolamo G, et al. 1992. AIDS-related complex treated by antiviral drugs and allogeneic bone marrow transplantation following conditioning protocol with busulphas, cyclophosphamide and cyclosporin. *Haematologica* 77:287–90
81. Contu L, LaNasa G, Arras M, et al. 1993. Allogeneic bone marrow transplantation combined with multiple anti-HIV-1 treatment in a case of AIDS. *Bone Marrow Transplant* 12:669–71
82. Gabarre J, Leblond V, Sutton L, et al. 1996. Autologous bone marrow transplantation in relapsed HIV-related non-Hodgkin's lymphoma. *Bone Marrow Transplant* 18:1195–97
83. Zaia J, Rossi J, Ito J, et al. 1998. One year results after autogolous stem cell transplantation using retrovirus-transduced peripheral blood progenitor cells in HIV-infected subjects. *Blood* 92:665a (Abstr.)
84. Krishnan A, Molina A, Zaia J, et al. 2001. Autologous stem cell transplantation for HIV-associated lymphoma. *Blood* 98:3857–59
85. Kohn DB, Bauer G, Rice CR, et al. 1999. A clinical trial of retroviral-mediated transfer of a rev-responsive element decoy gene into $CD34^+$ cells from the bone marrow of human immunodeficiency virus-1–infected children. *Blood* 94:368–71
86. Ratner L, Lee J, Tang S, et al. 2001. Chemotherapy for human immunodeficiency virus-associated non-Hodgkin's lymphoma in combination with highly active antiretroviral therapy. *J. Clin. Oncol.* 19(8):2171–78
87. Little RF, Yarchoan R, Wilson WH. 2000. Systemic chemotherapy for HIV-associated lymphoma in the era of highly active antiretroviral therapy. *Curr. Opin. Oncol.* 12:438–44
88. Chang Y, Cesarman E, Pessin MS, et al. 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* 266:1865–69
88a. Boshoff C, Chang Y. 2001. Kaposi's sarcoma-associated herpesvirus: a new DNA tumor virus. *Annu. Rev. Med.* 52: 453–70
89. Russo JJ, Bohenzky RA, Chien MC, et al. 1996. Nucleotide sequence of the Kaposi sarcoma–associated herpesvirus (HHV8). *Proc. Natl. Acad. Sci. USA* 93:14862–67
90. Ganem D. 1997. KSHV and Kaposi's sarcoma: the end of the beginning? *Cell* 91:157–60
91. Kedes DH, Operskalski E, Busch M, et al. 1996. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma–associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. *Nat. Med.* 2:918–24; erratum. 1996. *Nat. Med.* 2:1041
92. Min J, Katzenstein D. 1998. Detection of human herpesvirus 8 DNA in peripheral blood mononuclear cells of HIV+ subjects, with and without KS. *5th Conf. Retrovir. Oppor. Infect. 1998*, 160 (Abstr. 432)
93. Martin JN, Ganem DE, Osmond DH, et al. 1998. Sexual transmission and the natural history of human herpesvirus 8 infection. *N. Engl. J. Med.* 338:948–54
94. Chatlynne LG, Lapps W, Handy M, et al. 1998. Detection and titration of human herpesvirus-8-specific antibodies in sera from blood donors, acquired immunodeficiency syndrome patients, and Kaposi's

sarcoma patients using a whole virus enzyme-linked immunosorbent assay. *Blood* 92:53–58

95. Pauk J, Huang ML, Brodie SJ, et al. 2000. Mucosal shedding of human herpesvirus 8 in men. *N. Engl. J. Med.* 343:1369–77
96. Akula SM, Pramod NP, Wang FZ, Chandran B. 2002. Integrin alpha3beta1 (CD 49c/29) is a cellular receptor for Kaposi's sarcoma–associated herpesvirus (KSHV/HHV-8) entry into the target cells. *Cell* 108:407–19
97. Lagunoff M, Bechtel J, Venetsanakos E, et al. 2002. De novo infection and serial transmission of Kaposi's sarcoma–associated herpesvirus in cultured endothelial cells. *J. Virol.* 76:2440–48
98. Lee H, Veazey R, Williams K, et al. 1998. Deregulation of cell growth by the K1 gene of Kaposi's sarcoma–associated herpesvirus. *Nat. Med.* 4:435–40
99. Li M, Lee H, Guo J, et al. 1998. Kaposi's sarcoma–associated herpesvirus viral interferon regulatory factor. *J. Virol.* 72:5433–40
100. Zimring JC, Goodbourn S, Offermann MK. 1998. Human herpesvirus 8 encodes an interferon regulatory factor (IRF) homolog that represses IRF-1–mediated transcription. *J. Virol.* 72:701–7
101. Gao SJ, Boshoff C, Jayachandra S, et al. 1997. KSHV ORF K9 (vIRF) is an oncogene which inhibits the interferon signaling pathway. *Oncogene* 15:1979–85
102. Muralidhar S, Pumfery AM, Hassani M, et al. 1998. Identification of kaposin (open reading frame K12) as a human herpesvirus 8 (Kaposi's sarcoma–associated herpesvirus) transforming gene. *J. Virol.* 72:4980–88; erratum. 1999. *J. Virol.* 73:2568
103. Bais C, Santomasso B, Coso O, et al. 1998. G-protein–coupled receptor of Kaposi's sarcoma–associated herpesvirus is a viral oncogene and angiogenesis activator. *Nature* 391:86–89; erratum. 1998. *Nature* 392:210
104. Yang TY, Chen SC, Leach MW, et al. 2000. Transgenic expression of the chemokine receptor encoded by human herpesvirus 8 induces an angioproliferative disease resembling Kaposi's sarcoma. *J. Exp. Med.* 191:445–54
105. Martin DF, Kuppermann BD, Wolitz RA, et al. 1999. Oral ganciclovir for patients with cytomegalovirus retinitis treated with a ganciclovir implant. Roche Ganciclovir Study Group. *N. Engl. J. Med.* 340:1063–70
106. Mocroft A, Youle M, Gazzard B, et al. 1996. Anti-herpesvirus treatment and risk of Kaposi's sarcoma in HIV infection. Royal Free/Chelsea and Westminster Hospitals Collaborative Group. *AIDS* 10:1101–5
107. Glesby MJ, Hoover DR, Weng S, et al. 1996. Use of antiherpes drugs and the risk of Kaposi's sarcoma: data from the Multicenter AIDS Cohort Study. *J. Infect. Dis.* 173:1477–80
108. Robles R, Lugo D, Gee L, Jacobson MA. 1999. Effect of antiviral drugs used to treat cytomegalovirus end-organ disease on subsequent course of previously diagnosed Kaposi's sarcoma in patients with AIDS. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 20:34–38
109. Dairaghi DJ, Fan RA, McMaster BE, et al. 1999. HHV8-encoded vMIP-I selectively engages chemokine receptor CCR8. Agonist and antagonist profiles of viral chemokines. *J. Biol. Chem.* 274:21569–74
110. Endres MJ, Garlisi CG, Xiao H, et al. 1999. The Kaposi's sarcoma–related herpesvirus (KSHV)–encoded chemokine vMIP-I is a specific agonist for the CC chemokine receptor (CCR)8. *J. Exp. Med.* 189:1993–98
111. Aoki Y, Yarchoan R, Wyvill K, et al. 2001. Detection of viral interleukin-6 in Kaposi sarcoma–associated herpesvirus–linked disorders. *Blood* 97:2173–76
112. Murphy M, Armstrong D, Sepkowitz KA, et al. 1997. Regression of AIDS-related

Kaposi's sarcoma following treatment with an HIV-1 protease inhibitor. *AIDS* 11:261–62

113. Aboulafia D. 1998. Regression of AIDS-related pulmonary Kaposi's sarcoma after highly active antiretroviral therapy. *Mayo Clin. Proc.* 73:439–43
114. Coscoy L, Ganem D. 2000. Kaposi's sarcoma–associated herpesvirus encodes two proteins that block cell surface display of MHC class I chains by enhancing their endocytosis. *Proc. Natl. Acad. Sci. USA* 97:8051–56
115. Ishido S, Wang C, Lee BS, et al. 2000. Downregulation of major histocompatibility complex class I molecules by Kaposi's sarcoma–associated herpesvirus K3 and K5 proteins. *J. Virol.* 74:5300–9
116. Brander C, Suscovich T, Lee Y, et al. 2000. Impaired CTL recognition of cells latently infected with Kaposi's sarcoma–associated herpes virus. *J. Immunol.* 165:2077–83
117. Coscoy L, Ganem D. 2001. A viral protein that selectively downregulates ICAM-1 and B7-2 and modulates T cell costimulation. *J. Clin. Invest.* 107:1599–606
118. Harrington W Jr, Sieczkowski L, Sosa C, et al. 1997. Activation of HHV-8 by HIV-1 tat. *Lancet* 349:774–75
119. Ambrosino C, Ruocco MR, Chen X, et al. 1997. HIV-1 Tat induces the expression of the interleukin-6 (IL6) gene by binding to the IL6 leader RNA and by interacting with CAAT enhancer-binding protein beta (NF-IL6) transcription factors. *J. Biol. Chem.* 272:14883–92
120. Albini A, Soldi R, Giunciuglio D, et al. 1996. The angiogenesis induced by HIV-1 tat protein is mediated by the Flk-1/KDR receptor on vascular endothelial cells. *Nat. Med.* 2:1371–75
121. Cattelan A, Calabro M, Gasperini P, et al. 2001. Acquired immunodeficiency syndrome-related Kaposi's sarcoma regression after highly active antiretroviral therapy: biologic correlates of clinical outcome. *J. Natl. Cancer Inst. Monogr* 28:44–49
122. Dupont C, Vasseur E, Beauchet A, et al. 2000. Long-term efficacy on Kaposi's sarcoma of highly active antiretroviral therapy in a cohort of HIV-positive patients. CISIH 92. Centre d'Information et de Soins de l'Immunodeficience Humaine. *AIDS* 14:987–93
123. Krown SE, Testa MA, Huang J. 1997. AIDS-related Kaposi's sarcoma: prospective validation of the AIDS Clinical Trials Group staging classification. AIDS Clinical Trials Group Oncology Committee. *J. Clin. Oncol.* 15:3085–92
124. Bodsworth N. 1998. Topical 9–cis–retinoic acid (Panretin) gel as treatment of cutaneous AIDS-related Kaposi's sarcoma: interim results of an international, placebo-controlled trial (ALRT 1057-503). International Panretin KS Study Group. *Int. Conf. AIDS 1998* 12:317 (Abstr. 22277)
125. Gompels MM, Hill A, Jenkins P, et al. 1992. Kaposi's sarcoma in HIV infection treated with vincristine and bleomycin. *AIDS* 6:1175–80; erratum. 1992. *AIDS* 6:1410
126. Gill PS, Rarick MU, Espina B, et al. 1990. Advanced acquired immune deficiency syndrome–related Kaposi's sarcoma. Results of pilot studies using combination chemotherapy. *Cancer* 65:1074–78
127. Lee FC, Mitsuyasu RT. 1996. Chemotherapy of AIDS-related Kaposi's sarcoma. *Hematol. Oncol. Clin. N. Am.* 10:1051–68
128. Northfelt DW, Martin FJ, Working P, et al. 1996. Doxorubicin encapsulated in liposomes containing surface-bound polyethylene glycol: pharmacokinetics, tumor localization, and safety in patients with AIDS-related Kaposi's sarcoma. *J. Clin. Pharmacol.* 36:55–63
129. Northfelt DW, Dezube BJ, Thommes JA, et al. 1998. Pegylated-liposomal doxorubicin versus doxorubicin, bleomycin, and

vincristine in the treatment of AIDS-related Kaposi's sarcoma: results of a randomized phase III clinical trial. *J. Clin. Oncol.* 16:2445–51

130. Stewart S, Jablonowski H, Goebel FD, et al. 1998. Randomized comparative trial of pegylated liposomal doxorubicin versus bleomycin and vincristine in the treatment of AIDS-related Kaposi's sarcoma. International Pegylated Liposomal Doxorubicin Study Group. *J. Clin. Oncol.* 16:683–91
131. Gill PS, Wernz J, Scadden DT, et al. 1996. Randomized phase III trial of liposomal daunorubicin versus doxorubicin, bleomycin, and vincristine in AIDS-related Kaposi's sarcoma. *J. Clin. Oncol.* 14:2353–64
132. Welles L, Saville MW, Lietzau J, et al. 1998. Phase II trial with dose titration of paclitaxel for the therapy of human immunodeficiency virus–associated Kaposi's sarcoma. *J. Clin. Oncol.* 16:1112–21
133. Gill P, Tulpule A, Espina B, et al. 1999. Paclitaxel is safe and effective in the treatment of advanced AIDS-related Kaposi's sarcoma. *J. Clin. Oncol.* 17:1876–83
134. Little RF, Wyvill KM, Pluda JM, et al. 2000. Activity of thalidomide in AIDS-related Kaposi's sarcoma. *J. Clin. Oncol.* 18(13):2593–602
135. Dezube BJ, Von Roenn JH, Holden-Wiltse J, et al. 1998. Fumagillin analog in the treatment of Kaposi's sarcoma: a phase I AIDS Clinical Trial Group study. AIDS Clinical Trial Group No. 215 Team. *J. Clin. Oncol.* 16:1444–49

ELSEVIER

Brain Research 781 (1998) 244–251

BRAIN RESEARCH

Research report

Intracerebral HIV glycoprotein (gp120) enhances tumor metastasis via centrally released interleukin-1

Deborah M. Hodgson [a,*], Raz Yirmiya [d], Francesco Chiappelli [a,b], Anna N. Taylor [a,c]

[a] Dept. of Neurobiology and Brain Research Institute, School of Medicine, University of California, Los Angeles, CA 90095, USA
[b] Dept. of Neurobiology and Brain Research Institute, School of Dentistry, University of California, Los Angeles, CA 90095, USA
[c] West L.A. DVA Medical Center, Los Angeles, CA 90095, USA
[d] Dept. of Psychology, Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91905, Israel

Accepted 7 October 1997

Abstract

Infection with the human immunodeficiency virus (HIV) is associated with a high incidence of cancers. This relationship does not appear to be due to a direct effect of the virus, and may be mediated by neuroimmune interactions since the HIV glycoprotein, gp120, enters the brain soon after infection with HIV, and intracerebroventricular (i.c.v.) infusion of gp120 suppresses aspects of cellular and tumor immunity. It has been speculated that this suppression may be attributed to the release of interleukin-1 (IL-1) in the brain induced by gp120. Using an in vivo tumor model, we examined the effect of centrally administered gp120 on tumor metastasis and lung clearance of mammary adenocarcinoma (MADB106) tumor cells in rats, and the role played by brain IL-1 in mediating these effects. We demonstrate that central administration of gp120 (4 μg) significantly ($p < 0.05$) increased the retention of tumor cells in the lungs and significantly ($p < 0.02$) enhanced the development of tumor metastases. Central administration of IL-1β (10 ng) also significantly ($p < 0.05$) increased retention of tumor cells in the lungs. The effect of gp120 on lung retention of tumor cells was blocked by co-administration of α-melanocyte stimulating hormone (α-MSH, 20 ng), a hormone that blocks many of the biological effects of IL-1, or the IL-1 receptor antagonist (50 μg). Given that systemic administration of gp120 or IL-1β had no effect on the retention of tumor cells in the lungs, these findings indicate that gp120-induced secretion of IL-1 within the brain most likely mediates the effects of gp120 on tumor metastasis. These findings suggest a possible neuroimmune mechanism to account for the increased incidence and aggressiveness of tumors in HIV-infected patients. © 1998 Elsevier Science B.V.

Keywords: gp120; HIV; MADB106; IL-1; Tumor; Metastasis

1. Introduction

Human-immunodeficiency-virus (HIV) infection is associated with an increased incidence of malignant neoplasms. In addition to Kaposi's sarcoma and nonHodgkins lymphoma, which are particularly prevalent in AIDS patients, oral, rectal, testicular, and lung cancers have all been found to be associated with HIV infection [1,10,19,20,50]. In each instance the cancer is particularly aggressive and resistant to treatment [11,15]. The relationship between HIV infection and tumorigenesis appears to be indirect since HIV, in contrast to human oncoretroviruses, does not have in-vitro-transforming capacity nor is the HIV provirus found in any of the tumors associated with AIDS [15,19,50]. Tumor immunity is, however, compromised in HIV-infected individuals. Progression from HIV infection to AIDS is associated with a dramatic decline in the number and activity of natural killer (NK) cells, which are critically involved in the surveillance and early eradication of tumor cells [28,33,37]. The suppression of NK activity therefore may be related to the increased incidence of tumors in HIV infected individuals. Moreover, there is evidence to suggest that this immune suppression is mediated by central mechanisms activated by the entry of HIV into the brain [31,40–42]. On the basis of these findings we propose that the central actions of HIV and subsequent neuroimmune interactions may be of particular importance in understanding the increase in tumor incidence associated with HIV infection.

* Corresponding author. Dept. of Neurobiology University of California Los Angeles, Franz Hall Rm. 1282a, 405 Hilgard Ave., Los Angeles, CA 90095, USA. Fax: +1-310-8247471; E-mail: hodgson@psych.ucla.edu

syngeneic to the F344 rat, reliably colonizes to the lungs, forming well-defined surface metastases by 4 weeks post inoculation. Cells were maintained in 5% CO_2 at 37°C in monolayer cultures. They were grown in complete medium: RPMI 1640 media (Gibco, Grand Island, NY) supplemented with 10% heat-inactivated fetal bovine serum, L-glutamine (2 mM), non-essential amino acids (0.1 mM), sodium pyruvate (1 mM), and Gentamycin (0.01 mg/ml). Cells were trypsinized (0.25%) to remove from the flask. For DNA labeling, 0.4 μCi ^{125}IDUR (ICN Chemicals, Irvine, CA) per ml complete medium was added to the cell culture two days prior to harvesting. Prior to use, cells were separated from the flask, washed twice, and resuspended in PBS.

2.5. Assessment of tumor metastases and lung clearance of radiolabeled cells

Two hours after i.c.v. infusions, rats were lightly anaesthetized with halothane and 1×10^5 tumor cells in 0.5 ml PBS were injected into the tail vein. Animals were returned to their home cages and 4 weeks later rats were euthanized. At this time lungs were removed, fixed in Bouins solution for 24 h, then transferred into ethanol. Surface metastases were counted by two independent observers, blind to group membership, and skilled in tumor identification. Previous research by this and other laboratories has established that colonization of the lungs and growth of MADB106 tumors is stable and predictable across experiments [3–6]. The most reliable time point to assess metastases is at 3–4 weeks post inoculation. At 2 weeks post inoculation metastases resemble white spheres up to 0.5 mm^2 and are only slightly raised above the lung surface. At 4 weeks post inoculation, however, surface metastases are 1–8 mm^3, mushroom-shaped, and distinctly separated and raised above the lung surface, which allows for accurate enumeration of tumor metastases. For the assessment of lung clearance, rats were lightly anaesthetized with halothane and 1×10^5 radiolabeled tumor cells in 0.5 ml PBS were injected into the tail vein. Animals were returned to their home cages with food and water freely available. Six hours later, animals were euthanized, lungs were removed, and radioactivity was assessed using a gamma counter. Percent radioactivity was the amount of radioactivity detected in the lungs compared to the amount of radioactivity present in the labelled MADB106 tumor cells prior to inoculation.

3. Experimental procedures

3.1. Expt. 1A: Effect of i.c.v. administration of gp120 on the metastasis of MADB106 tumor cells

One week after cannula implantation surgery, animals were randomly divided into two groups: gp120 ($n = 25$) and PBS ($n = 15$). One group of animals (gp120) was infused i.c.v. (10 μl/min) with gp120 (4 μg gp120/10 μl PBS), the other group was infused with an equivalent volume of the vehicle (PBS). Animals were returned to their home cages and 2 h later were lightly anaesthetized with halothane and inoculated via the tail vein with MADB106 tumor cells (1×10^5/0.5 ml PBS). Four weeks after inoculation animals were euthanized with halothane and lungs were obtained for enumeration of tumor metastases.

3.2. Expt. 1B: Effect of i.c.v. administration of gp120 on lung clearance of MADB106 tumor cells

One week after cannulation surgery animals were randomly divided into two groups ($n = 15$/group). One group of animals (gp120) was infused i.c.v. (10 μl/min) with gp120 (4 μg/10 μl PBS), the other group (PBS) was infused with an equivalent volume of the vehicle. Animals were returned to their home cages and 2 h later were lightly anaesthetized with halothane and inoculated via the tail vein with radiolabeled [^{125}IDUR] MADB106 tumor cells (1×10^5/0.5 ml PBS). Six hours later, animals were euthanized with halothane and lungs were obtained for assessment of radioactivity.

3.3. Expt. 2: Effect of i.c.v. administration of IL-1β on lung clearance of MADB106 tumor cells

The same procedure as in Expt. 1B was used to examine the effects of IL-1β on lung clearance except that one group of animals (IL-1β) was infused with IL-1β (10 ng/10 μl), and a second group (PBS) was infused with the equivalent volume of the vehicle, PBS ($n = 10$/group). Two hours later, the animals were lightly anaesthetized with halothane and inoculated via the tail vein with radiolabeled [^{125}IDUR] MADB106 tumor cells (1×10^5/0.5 ml PBS). Six hours later, animals were euthanized with halothane and lungs were obtained for assessment of radioactivity.

3.4. Expt. 3A: Effect of i.c.v. co-administration of gp120 and α-MSH on lung clearance of MADB106 tumor cells.

One week after cannula implantation surgery animals were randomly divided into four groups ($n = 8$/group): PBS/PBS, PBS/gp120, α-MSH/PBS, α-MSH/gp120. Animals were infused i.c.v. with either α-MSH (20 ng/10 μl PBS), or vehicle (10 μl/PBS), followed by a second infusion of either gp120 (4 μg/10 μl PBS), or vehicle (10 μl PBS). Animals were returned to their home cages and 2 h later, were lightly anaesthetized with halothane and inoculated via the tail vein with radiolabeled [^{125}IDUR] MADB106 tumor cells (1×10^5/0.5 ml PBS). Six hours later, animals were euthanized and lungs were obtained for assessment of radioactivity.

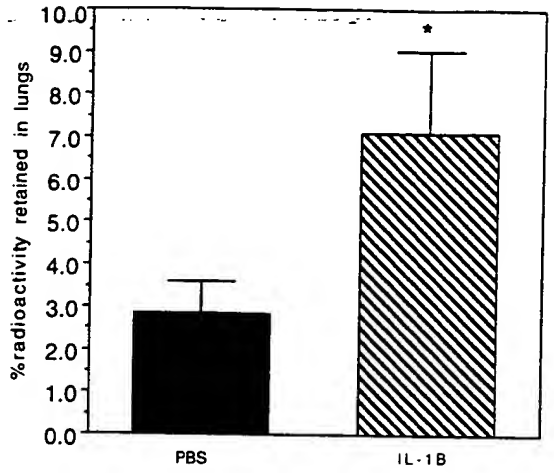


Fig. 3. The effect of IL-1β administration on the percentage of radioactivity (mean ± S.E.M.) retained in the lungs 6 h after inoculation with radiolabeled MADB106 tumor cells. $^{*}p < 0.05$ compared with the PBS control group.

to significantly increase the retention of tumor cells in the lungs compared to vehicle treated animals [$t(14) = 2.45$, $p = 0.028$].

4.3. Expt. 3A and 3B: effect of i.c.v. co-administration of gp120 and either α-MSH or IL-Ira on lung clearance of MADB106 tumor cells

Fig. 4 illustrates the effect of co-administration of gp120 with α-MSH on lung clearance of radiolabeled MADB106 tumor cells. A two-way ANOVA indicated a significant

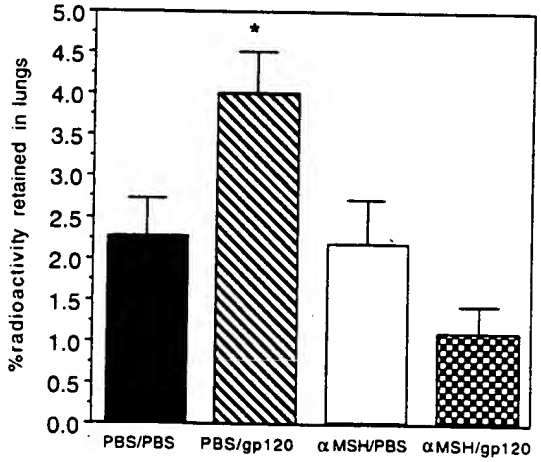


Fig. 4. The effect of co-administration of gp120 with α-MSH on the percentage of radioactivity (mean ± S.E.M.) retained in the lungs 6 h after inoculation with radiolabeled MADB106 tumor cells. Group PBS/PBS was i.c.v. infused with PBS(10 μl) followed by PBS(10 μl). Group PBS/gp120 was infused with PBS (10 μl) followed by gp120 (4 μg/10 μl PBS). Group α-MSH/PBS was infused with α-MSH (20 ng/10 μl PBS), followed by the vehicle (10 μl/PBS). Group α-MSH/gp120 was infused with α-MSH (20 ng/10 μl PBS), followed by gp120 (4 μg/10 μl PBS). $^{*}p < 0.05$ compared to the PBS, α-MSH, and combined α-MSH/gp120 groups.

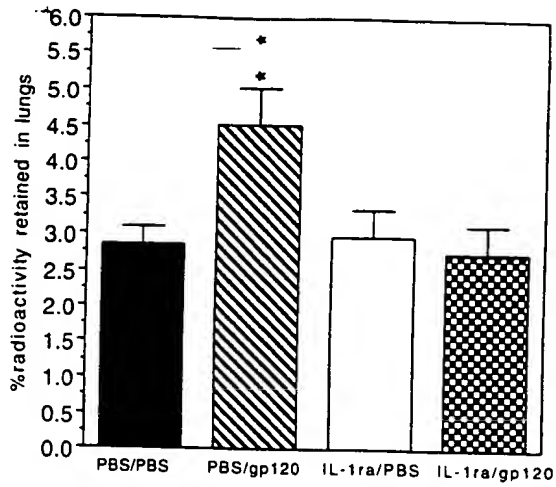


Fig. 5. The effect of co-administration of gp120 with IL-1 ra on the percentage of radioactivity (mean ± S.E.M.) retained in the lungs 6 h after inoculation with radiolabeled MADB 106 tumor cells. Group PBS/PBS was i.c.v. infused with PBS(10 μl) followed by PBS (10 μl). Group PBS/gp120 was infused with PBS (10 μl) followed by gp120 (4 μg/10 μl PBS). Group IL-1 ra/PBS was infused with IL-1 ra (50 μg/10 μl PBS), followed by the vehicle (10 μl/PBS). Group IL-1 ra/gp120 was infused with IL-1 ra (50 μg/10 μl PBS), followed by gp120 (4 μg/10 μl PBS). $^{**}p < 0.01$ compared to the PBS, IL-1 ra, and the combined IL-1 ra/PBS groups.

interaction between the first (α-MSH/PBS) and second injection (PBS/gp120) [$F_{1,24} = 6.071$, $p = 0.02$]. Bonferroni comparisons indicated that gp120 significantly ($p < 0.05$) increased the retention of MADB106 tumor cells in the lungs compared to the vehicle injection. The effect of gp120 was blocked by co-infusion of α-MSH ($p < 0.01$), and α-MSH had no effect itself. Fig. 5 illustrates the effect of co-administration of gp120 with IL-1 ra on lung clearance of radiolabeled MADB106 tumor cells. A two-way ANOVA indicated a significant interaction between the first (IL-1 ra/PBS) and second injection (PBS/gp120) ($F_{1,71} = 5.371$, $p = 0.02$). Bonferroni comparisons indicate that gp120 significantly ($p < 0.01$) increased retention of MADB106 tumor cells in the lungs, and this was blocked by co-infusion of IL-1 ra ($p < 0.01$). IL-1 ra had no effect itself.

4.4. Expt. 4: effect of systemic administration of gp120 or IL-1β on lung clearance of MADB106 tumor cells

Data analysis revealed there were no significant differences between the three groups. There was no effect of either gp120 or IL-1β on lung clearance of labeled MADB106 tumor cells when compared to vehicle-treated controls.

5. Discussion

The present study demonstrates that decreased lung clearance and enhanced lung colonization of MADB106

mals in the absence of adrenal hormones [42]. A neural pathway has been implicated given that blocking neural transmission at sympathetic ganglia partially attenuates the IL-1-induced suppression of NK cell activity [45]. The sympathetic nervous system seems to be of particular relevance given that centrally administered IL-1 increases splenic sympathetic activity [18]. Electrical stimulation of the splenic nerve also reduces NK cell activity, and this effect is blocked by pretreatment with β-receptor antagonists [22]. Furthermore, activation of the β-2 adrenergic receptors mediates the suppressive effects of acute stress on NK cell activity and results in increased metastatic spread of MADB106 tumor cells in the rat [6]. Future research should address the role of neural and hormonal mechanisms in mediating the promotion of tumor metastasis by gp120.

6. Conclusion

In conclusion, the present study demonstrates that entry of the HIV gp120 into the brain enhances tumor metastasis and this appears to be mediated by the central release of IL-1. The inhibition of lung clearance and enhancement of tumor metastasis by gp120 is most likely mediated by the suppression of NK- and T cell-mediated immunity in the periphery, possibly via IL-1-induced activation of the HPA axis and the sympathetic nervous system. These findings may provide a mechanism to account for the finding that HIV infection is associated with a high incidence of malignancy. Although significant progress is being made with aggressive combined chemotherapy in the treatment of AIDS, in cases where the disease is complicated by the presence of malignancy, the prognosis remains exceptionally poor [32,36]. Treatment options are limited because the poor immunological status of AIDS patients renders them intolerant to standard anti-tumor therapy [10]. Thus, there is a need to develop new approaches for effective treatment of HIV-related malignancies. Attenuation of the immunosuppressive and tumor enhancing effects of gp120 by selective pharmacological blockade of the central effects of IL-1 may provide such an approach.

Acknowledgements

This research was supported by The Norman Cousins Program in Psychoneuroimmunology at UCLA (D.M.H.), Department of Veterans Affairs Medical Research Service, NIH/NIAAA AA09850, and the US–Israel Binational Science Foundation 94-00062 (R.Y and A.N.T). We also acknowledge the generous donation of IL-1 from the National Cancer Institute (Biological Response Modifiers Program, NIH/NCI) and IL-1 receptor antagonist from AMGEN (Boulder, CO, USA). The expert technical assistance of Mr. Ngy Heng and Khai Nguyen was also greatly appreciated.

References

[1] D.M. Bacr, E.T. Anderson, L.S. Wilkinson, Acquired immune deficiency syndrome in homosexual men with Hodgkin's disease, Am. J. Med. 80 (1986) 738–740.
[2] T. Barlozzari, J. Leonhardt, R.H. Wiltrout, R.B. Herberman, C.W. Reynolds, Direct evidence for the role of LGL in the inhibition of experimental tumor metastases, J. Immunol. 134 (1985) 2783–2789.
[3] S. Ben-Eliyahu, G.G. Page, In vivo assessment of natural killer cell activity in rats, Prog. Neuroendocrinol. 5 (1992) 199–214.
[4] S. Ben-Eliyahu, G.G. Page, R. Yirmiya, A.N. Taylor, Acute alcohol intoxication suppresses natural killer cell activity and promotes tumor metastasis, Nature Med. 2 (1996) 457–460.
[5] S. Ben-Eliyahu, R. Yirmiya, J. Liebeskind, A.N. Taylor, R.P. Gale, Stress increases metastatic spread of a mammary tumor in rats: evidence for mediation by the immune system, Brain Behav. Immun. 5 (1991) 193–205.
[6] S. Ben-Eliyahu, R. Yirmiya, G. Page, H. Weiner, A. Tan, A.N. Taylor, J.C. Liebeskind, Stress-induced sympathetic activation suppresses blood natural killer cytotoxicity and increases metastatic spread in rats: mediation by adrenal epinephrine, Soc. Neurosci. Abstr., 17 (1991) 327, 165.*
[7] F. Berkenbosch, J. VanOers, A. DelRey, F. Tilders, H. Besedovsky, Corticotropin-releasing-factor-producing neurons in the rat activated by interleukin-1, Science 238 (1987) 524–526.
[8] H. Besedovsky, A.D. Del Rey, E. Sorkin, C.A. Dinarello, Immunological feedback between IL-1 and glucocorticoid hormones, Science 233 (1986) 652–654.
[9] R.S. Chirivi, A. Garofolo, I.M. Padura, A. Mantovani, R. Giavazzi, Interleukin-1 receptor antagonist inhibits the augmentation of metastasis induced by interleukin-1 or lipopolysaccaride in a melanoma/nude mouse system, Cancer Res. 53 (1993) 5051–5054.
[10] E.F. Di Carlo, J.B. Amberson, C.E. Metroka, P. Ballard, A. Moore, J.A. Mouradian, Malignant lymphomas and the acquired immunodeficiency syndrome, Arch. Pathol. Lab. Med. 110 (1986) 1012–1016.
[11] D. Errante, E. Vaccher, U. Tirelli et al., Management of AIDS and its neoplastic complications, Eur. J. Cancer 27 (1991) 380–389.
[12] R. Evans, M. Fong, J. Fuller, S. Kamdar, J. Meyerhardt, G. Strassmann, Tumor cell IL-6 gene expression is regulated by IL-1 α/β and TNFα: proposed feedback mechanisms induced by the interaction of tumor cells and macrophages, J. Leukoc. Biol. 52 (1992) 463–468.
[13] J.K. Ferdynand, E.G. Engelman, Immune regulation: a critical link between NK cells and CTLs, Immunol. Today 17 (1996) 174–176.
[14] S. Gillis, G.R. Crabtree, K.A. Smith, Glucocorticoid-induced inhibition of T-cell growth factor production: 1. The effect on mitogen-induced lymphocyte proliferation, J. Immunol. 123 (1979) 1624–1638.
[15] E.M. Hersch, T.P. Miller, Malignant lymphomas in patients with human immunodeficiency virus infection, Recent Results Cancer Res. 112 (1988) 54–62.
[16] D.M. Hodgson, A.N. Taylor, F. Chiappelli, M. Kung, N.S. Morrow, Impact of acute food restriction on the colonization of MADB106 tumor cells to the lungs of Fischer 344 rats, Neuroimmunomod. 3 (1996) 371–381.
[17] F. Homo-Delarche, F. Fitzpatrick, N. Christeff, E.A. Nunez, J.F. Bach, M. Dardenne, Sex steroids, glucocorticoids, stress and autoimmunity, J. Steroid Biochem. Mol. Biol. 40 (1991) 619–637.
[18] T. Ichijo, T. Katafuchi, T. Hori, Central interleukin-1β enhances splenic sympathetic nerve activity in rats, Brain Res. Bull. 34 (1994) 547–553.
[19] H.L. Ioachim, The opportunistic tumors of immune deficiency, Adv. Cancer. Res. 54 (1990) 301–317.

EXHIBIT A
Docket No.: PF-0636 RCE
USSN: 09/831,458

SeqServer®
biology in silico

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- GSEQ:AAR32188
- 3344986CD1

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: GSEQ_AAR32188      404 aa
Sequence 2: 3344986CD1         325 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  84
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:3942
Alignment Score 1706
CLUSTAL-Alignment file created  [baa2EaaTO.aln]
CLUSTAL W (1.7) multiple sequence alignment


GSEQ_AAR32188      MSDSKEPRLQQLGLLEEEQLRGLGFRQTRGYKSLAGCLGHGPLVLQLLSFTLLAG----L
3344986CD1         MSDSKEPRVQQLGLL--------------------GCLGHGALVLQLLSFMLLAGVLVAI
                   ********:******                     ******.******** ****    :

GSEQ_AAR32188      LVQVSKVPSSISQEQSRQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE
3344986CD1         LVQVSKVPSSLSQEQSEQDAIYQNLTQLKAAVGELSEKSKLQEIYQELTQLKAAVGELPE
                   **********:*****.*******************************************

GSEQ_AAR32188      KSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTWLKAAVGELPEKSKMQEIYQELTRL
3344986CD1         KSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTRLKAAVGELPEKSKLQEIYQELTRL
                   ********************************** ************:**********

GSEQ_AAR32188      KAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQEEIYQELTRLKAAVGELPEKSKQQ
3344986CD1         KAAVGELPD---Q-------------------------------------------SKQQ
                   ********:   *                                             ****

GSEQ_AAR32188      EIYQELTQLKAAVERLCHPCPWECTFFQGNCYFMSNSQRNWHDSITACKEVGAQLVVIKS
3344986CD1         QIYQELTDLKTAFERLCRHCPKDWTFFQGNCYFMSNSQRNWHDSVTACQEVRAQLVVIKT
                   :******:**:*.****: ** : ********************:***:** *******:

GSEQ_AAR32188      AEEQNFLQLQSSRSNRFTWMGLSDLNQEGTWQWVDGSPLLPSFKQYWNRGEPNNVGEEDC
3344986CD1         AEEQNFLQLQTSRSNRFSWMGLSDLNQEGTWQWVDGSPLSPSFQRYWNSGEPNNSGNEDC
```

```
                    **********:******:********************** ***::*** ***** *:***

GSEQ_AAR32188       AEFSGNGWNDDKCNLAKFWICKKSAASCSRDEEQFLSPAPATPNPPPA
3344986CD1          AEFSGSGWNDNRCDVDNYWICKKPAA-CFRDE----------------
                    *****.****::*:: ::******.** * ***
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

EXHIBIT B
Docket No.: PF-0636 RCE
USSN: 09/831,458

SeqServer®
biology in silico

BLAST2 Search Results

 SeqServer Version 4.6 Jan 2002

Program: blastp
Sequence ID(s):
3344986CD1 vs. genpept131

NCBI-BLASTP 2.0.10 [Aug-26-1999]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= 3344986CD1
(325 letters)

Database: genpept131
1,135,942 sequences; 348,344,575 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
g13383470 L-SIGN [Homo sapiens]	635	0.0
g15383606 mDC-SIGN2 type I isoform [Homo sapiens]	621	e-177
g12084795 probable mannose binding C-type lectin DC-SIGNR [Ho	621	e-177
g12084797 probable mannose binding C-type lectin DC-SIGNR [Ho	617	e-175
g15383614 sDC-SIGN2 type I isoform [Homo sapiens]	583	e-165
g8572543 membrane-associated lectin type-C [Homo sapiens]	532	e-150
g17049084 unnamed protein product [Homo sapiens]	532	e-150
g15281073 mDC-SIGN1A type I isoform [Homo sapiens]	532	e-150
g13383468 DC-SIGN [Homo sapiens]	532	e-150
g10179610 probable mannose-binding C-type lectin DC-SIGN [Hom	532	e-150

>g13383470 L-SIGN [Homo sapiens]
Length = 376

Score = 635 bits (1619), Expect = 0.0
Identities = 324/376 (86%), Positives = 325/376 (86%), Gaps = 51/376 (13%)

```
Query: 1   MSDSKEPRVQQLGLL----------------------------GCLGHGALVLQLLSFML 32
            MSDSKEPRVQQLGLL                            GCLGHGALVLQLLSFML
Sbjct: 1   MSDSKEPRVQQLGLLEEDPTTSGIRLFPRDFQFQQIHGHKSSTGCLGHGALVLQLLSFML 60
```

```
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /data/isb2k/blastdb/Pfam72/Pfam72
Sequence file:            /u/legal/jennyb/pf636.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:  3344986CD1

Scores for sequence family classification (score includes all domains):
Model         Description                                   Score    E-value  N
--------      -----------                                   -----    ------- ---
lectin_c      Lectin C-type domain                          139.5    5.9e-38   1
Ribosomal_L29 Ribosomal L29 protein                         -15.3        9.1   1

Parsed for domains:
Model         Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
--------      ------- ----- -----    ----- -----       -----  -------
Ribosomal_L29   1/1      85   152 ..     1    64 []    -15.3      9.1
lectin_c        1/1     211   317 ..     1   125 []    139.5  5.9e-38

Alignments of top-scoring domains:
Ribosomal_L29: domain 1 of 1, from 85 to 152: score -15.3, E = 9.1
                   *->akELRelsde..EL..eeeleelKrELfeLRAfqaAtGqLenPhrIk
                      ++EL +l+ + +EL+++ +l e+ +EL  L+   aA+G+L   +++
  3344986CD1    85    YQELTQLKAAvgELpeKSKLQEIYQELTRLK---AAVGELPEKSKLQ 128

                   evRkrIARilTv...lnErklsae<-*
                   e+ +++ R++ + ++l E+ + +e
  3344986CD1   129 EIYQELTRLKAAvgeLPEKSKLQE    152

lectin_c: domain 1 of 1, from 211 to 317: score 139.5, E = 5.9e-38
                   *->esktWaeAelaCqkegghAHLvsIqsaeEqsfvvafltsltkksnty
                      ++++W+++ +aCq+  ++  Lv+I  aeEq    +fl+  t++sn
  3344986CD1   211    SQRNWHDSVTACQEVRAQ--LVVIKTAEEQ----NFLQLQTSRSNRF 251

                   aWIGLtdintegtwvwegwetdgspvnyt..enWapgePnnrgnhGgnEd
                    W+GL+d n+egtw+w    +dgsp++ + +++W++gePnn gn    Ed
  3344986CD1   252 SWMGLSDLNQEGTWQW----VDGSPLSPSfqRYWNSGEPNNSGN----ED 293

                   Cveiytdtdflaa GkWnDepCdsklpyvCef<-*
                   C+e++++       WnD+ Cd+ + ++C++
  3344986CD1   294 CAEFSGS------GWNDNRCDVDNYWICKK    317
```

```
1   MSDSKEPRVQQLGLL----------------  SEQIDNO_12
1   MSDSKEPRLQQLGLLEEEQL---------RG  GENESEQ_AAR32188
1   MSDSKEPRLQQLGLLEEEQL---------RG  g8572543
1   MSDSKEPRVQQLGLLEEDPTTSGIRLFPRD  g13383470

16  ----------------GCLGHGALVLQLLSFML  SEQIDNO_12
23  LGFRQTRGYKSLAGCLGHGPLVLQLLSFTL  GENESEQ_AAR32188
23  LGFRQTRGYKSLAGCLGHGPLVLQLLSFTL  g8572543
31  FQFQQIHGHKSSTGCLGHGALVLQLLSFML  g13383470

33  LAGVLVAILVQVSKVPSSLSQEQSEQDAIY  SEQIDNO_12
53  LAGLLV----QVSKVPSSISQEQSRQDAIY  GENESEQ_AAR32188
53  LAGLLV----QVSKVPSSISQEQSRQDAIY  g8572543
61  LAGVLVAILVQVSKVPSSLSQEQSEQDAIY  g13383470

63  QNLTQLKAAVGELSEKSKLQEIYQELTQLK  SEQIDNO_12
79  QNLTQLKAAVGELSEKSKLQEIYQELTQLK  GENESEQ_AAR32188
79  QNLTQLKAAVGELSEKSKLQEIYQELTQLK  g8572543
91  QNLTQLKAAVGELSEKSKLQEIYQELTQLK  g13383470

93  AAVGELPEKSKLQEIYQELTRLKAAVGELP  SEQIDNO_12
109 AAVGELPEKSKLQEIYQELTRLKAAVGELP  GENESEQ_AAR32188
109 AAVGELPEKSKLQEIYQELTRLKAAVGELP  g8572543
121 AAVGELPEKSKLQEIYQELTRLKAAVGELP  g13383470

123 EKSKLQEIYQELTRLKAAVGELPEKSKLQE  SEQIDNO_12
139 EKSKLQEIYQELTWLKAAVGELPEKSKMQE  GENESEQ_AAR32188
139 EKSKLQEIYQELTWLKAAVGELPEKSKMQE  g8572543
151 EKSKLQEIYQELTRLKAAVGELPEKSKLQE  g13383470
```